GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:54:18 ; Search time 351.297 Seconds
(without alignments)
5302.013 Million cell updates/sec

Title: US-09-990-099-23
Perfect score: 64
Sequence: 1 caaaaaaaagtccgctgata.....cccatttttacatcgtagcc 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	64	100.0	720	1	AF407018	AF407018 Escherich
2	64	100.0	2886	1	ECOPROTS	M94104 Escherichia
c 3	64	100.0	10120	1	AE000446	AE000446 Escherich
c 4	64	100.0	12642	1	AE005600	AE005600 Escherich
c 5	64	100.0	136254	1	ECOUW82	L10328 E. coli; th
6	64	100.0	221285	2	AC023611	AC023611 Mus muscu
c 7	64	100.0	252977	2	AC073670	AC073670 Mus muscu
c 8	64	100.0	256373	2	AC020870	AC020870 Mus muscu
9	64	100.0	272545	2	AC090533	AC090533 Mus muscu
c 10	64	100.0	280900	1	AP002566	AP002566 Escherich
11	39.6	61.9	245050	1	AL627280	AL627280 Salmonell
c 12	38	59.4	22492	1	AE008877	AE008877 Salmonell
c 13	28.8	45.0	25869	9	AC002459	AC002459 Homo sapi
c 14	27.4	42.8	66873	2	AC129514	AC129514 Homo sapi
15	27.4	42.8	132288	9	AC099848	AC099848 Homo sapi
c 16	27.4	42.8	155935	9	AC006442	AC006442 Homo sapi
17	27.4	42.8	157028	9	AL139000	AL139000 Human DNA
c 18	27.4	42.8	166937	9	AC025679	AC025679 Homo sapi
19	27.4	42.8	178563	2	AP002419	AP002419 Homo sapi
c 20	27.4	42.8	187559	2	AC119816	AC119816 Mus muscu
21	27.4	42.8	191211	2	AC131053	AC131053 Homo sapi
22	27.4	42.8	317463	2	AC091254	AC091254 Mus muscu
23	27	42.2	98828	9	HSJ388E23	AL049552 Human DNA
c 24	27	42.2	130110	9	AL731777	AL731777 Human DNA
25	27	42.2	135968	2	AC021707	AC021707 Homo sapi
26	27	42.2	141854	2	AC021595	AC021595 Homo sapi
27	27	42.2	152449	2	AC080175	AC080175 Homo sapi
28	27	42.2	200175	2	AC096004	AC096004 Rattus no
29	26.8	41.9	22516	3	CEC50C10	Z72505 Caenorhabdi
30	26.8	41.9	90347	5	AL714030	AL714030 Zebrafish
31	26.8	41.9	145634	9	AC096716	AC096716 Homo sapi
c 32	26.8	41.9	196311	9	CNS01DVP	AL135999 Human chr
c 33	26.8	41.9	200598	2	AC012505	AC012505 Homo sapi
34	26.6	41.6	30274	3	CEF45H11	Z78420 Caenorhabdi
c 35	26.6	41.6	82047	3	CEY67A10A	AL117206 Caenorhab
36	26.6	41.6	90018	2	AC017853	AC017853 Drosophil
c 37	26.6	41.6	153439	8	AP003313	AP003313 Oryza sat
38	26.6	41.6	159555	2	AL591710	AL591710 Homo sapi
c 39	26.6	41.6	162591	9	AL391839	AL391839 Human DNA
c 40	26.6	41.6	167026	3	AC010011	AC010011 Drosophil
41	26.6	41.6	182567	2	AC102472	AC102472 Mus muscu
42	26.6	41.6	193902	2	AL591112	AL591112 Homo sapi
c 43	26.6	41.6	195349	2	AC006705	AC006705 Caenorhab
c 44	26.6	41.6	195431	2	AC130204	AC130204 Mus muscu
45	26.6	41.6	195879	2	AL807248	AL807248 Mus muscu

ALIGNMENTS

RESULT 1
AF407018/c
LOCUS AF407018 720 bp DNA linear BCT 10-SEP-2001
DEFINITION Escherichia coli inner membrane protein (ecfI) gene, complete cds.
ACCESSION AF407018
VERSION AF407018.1 GI:15529642
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 720)
AUTHORS Dartigalongue,C., Missiakas,D. and Raina,S.
TITLE Characterization of the Escherichia coli sigma E regulon
JOURNAL J. Biol. Chem. 276 (24), 20866-20875 (2001)

```
  MEDLINE   21293005
   PUBMED   11274153
REFERENCE   2  (bases 1 to 720)
  AUTHORS   Dartigalongue,C., Missiakas,D. and Raina,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Biochimie Medicale, CMU, 1 Rue Michel
            Servet, Geneva 1211, Switzerland
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            142. .>694
                     /gene="ecfI"
     promoter        142. .177
                     /gene="ecfI"
                     /note="factor sigma24"
     mRNA            178. .>694
                     /gene="ecfI"
                     /product="inner membrane protein"
     CDS             287. .694
                     /gene="ecfI"
                     /note="EcfI; under heat shock transcriptional control of
                     RpoE"
                     /codon_start=1
                     /transl_table=11
                     /product="inner membrane protein"
                     /protein_id="AAL01385.1"
                     /db_xref="GI:15529643"
                     /translation="MMMLFSARDRYNSSGQHPRSTDEFIMIRNVLLAFMICSGMTLLG
                     GCSSVMSHTGGKEGTYPGTRASATMIGDDETNWGTKSLAILDMPFTAVMDTLLLPWDV
                     FRKDSSVRSRVEKSEANAQATNAVIPPARMPDN"
BASE COUNT      207 a    169 c    183 g    161 t
ORIGIN

  Query Match           100.0%;  Score 64;  DB 1;  Length 720;
  Best Local Similarity  100.0%;  Pred. No. 6.4e-11;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    210 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 151

Qy     61 AGCC 64
          ||||
Db    150 AGCC 147


RESULT 2
ECOPROTS
LOCUS       ECOPROTS                2886 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Escherichia coli ibpA and ibpB genes, complete cds.
ACCESSION   M94104
VERSION     M94104.1  GI:147368
KEYWORDS    ibpA gene; ibpB gene.
SOURCE      Escherichia coli (sub_strain W3110, strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 2886)
  AUTHORS   Allen,S.P., Polazzi,J.O., Gierse,J.K. and Easton,A.M.
  TITLE     Two novel heat shock genes encoding proteins produced in response
            to heterologous protein expression in Escherichia coli
  JOURNAL   J. Bacteriol. 174 (21), 6938-6947 (1992)
  MEDLINE   93015757
   PUBMED   1356969
FEATURES             Location/Qualifiers
     source          1. .2886
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /sub_strain="W3110"
                     /db_xref="taxon:562"
     gene            539. .1083
                     /gene="ibpA"
     -35_signal      539. .545
                     /gene="ibpA"
                     /note="putative"
     -10_signal      560. .566
                     /gene="ibpA"
                     /note="putative"
     RBS             656. .660
                     /gene="ibpA"
                     /note="putative"
     CDS             670. .1083
                     /gene="ibpA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA24424.1"
                     /db_xref="GI:147369"
                     /translation="MRNFDLSPLYRSAIGFDRLFNHLENNQSQSNGGYPPYNVELVDE
                     NHYRIAIAVAGFAESELEITAQDNLLVVKGAHADEQKERTYLYQGIAERNFERKFQLA
                     ENIHVRGANLVNGLLYIDLERVIPEAKKPRRIEIN"
     gene            1181. .1623
                     /gene="ibpB"
     RBS             1181. .1186
                     /gene="ibpB"
                     /note="putative"
     CDS             1195. .1623
                     /gene="ibpB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA24425.1"
                     /db_xref="GI:147370"
                     /translation="MRNFDLSPLMRQWIGFDKLANALQNAGESQSFPPYNIEKSDDNH
                     YRITLALAGFRQEDLEIQLEGTRLSVKGTPEQPKEEKKWLHQGLMNQPFSLSFTLAEN
                     MEVSGATFVNGLLHIDLIRNEPEPIAAQRIAISERPALNS"
     terminator      1639. .1666
                     /note="putative"
     stem_loop       1639. .1659
                     /note="putative"
     -35_signal      1769. .1774
                     /note="putative"
     -10_signal      1791. .1796
                     /note="putative"
     RBS             1805. .1809
                     /note="putative"
     CDS             1819. .2862
                     /note="ORF; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA24426.1"
                     /db_xref="GI:147371"
                     /translation="MSDIALTVSILALVAVVGLFIGNVKFRGIGLGIGGVLFGGIIVG
                     HFVSQAGMTLSSDMLHVIQEFGLILFVYTIGIQVGPGFFASLRVSGLRLNLFAVLIVI
                     IGGLVTAILHKLFDIPLPVVLGIFSGAVTNTPALGAGQQILRDLGTPMEMVDQMGMSY
                     AMAYPFGICGILFTMWMLRVIFRVNVETEAQQHESSRTNGGALIKTINIRVENPNLHD
                     LAIKDVPILNGDKIICSRLKREETLKVPSPDTIIQLGDLLHLVGQPADLHNAQLVIGQ
                     EVDTSLSTKGTDLRVERVVVTNENVLGKRIRDLHFKERYDVVISRLNRAGVELVASGD
                     ISLQRTDRNAKLL"
BASE COUNT      693 a    701 c    722 g    770 t
ORIGIN

  Query Match           100.0%;  Score 64;  DB 1;  Length 2886;
  Best Local Similarity  100.0%;  Pred. No. 5e-11;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    516 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 575

Qy     61 AGCC 64
          ||||
Db    576 AGCC 579
```

RESULT 3
AE000446/c
LOCUS AE000446 10120 bp DNA linear BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 336 of 400 of the complete genome.
ACCESSION AE000446 U00096
VERSION AE000446.1 GI:2367261
KEYWORDS .
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10120)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 10120)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 10120)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 10120)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT On Sep 9, 1997 this sequence version replaced gi:1790121.
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES Location/Qualifiers
source 1. .10120
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
promoter complement(16. .46)
/note="factor Sigma70; predicted +1 start at 3864562"
promoter complement(27. .58)
/note="factor Sigma70; predicted +1 start at 3864573"
gene complement(83. .496)
/gene="ibpA"
/note="b3687"
CDS complement(83. .496)
/gene="ibpA"
/function="factor; Adaptations, atypical conditions"
/note="f137; heat shock inducible; alternate gene name hslT"
/codon_start=1
/transl_table=11
/product="heat shock protein"
/protein_id="AAC76710.1"
/db_xref="GI:1790122"
/translation="MRNFDLSPLYRSAIGFDRLFNHLENNQSQSNGGYPPYNVELVDE NHYRIAIAVAGFAESELEITAQDNLLVVKGAHADEQKERTYLYQGIAERNFERKFQLA ENIHVRGANLVNGLLYIDLERVIPEAKKPRRIEIN"
promoter complement(146. .163)
/gene="ibpA"
/note="factor Sigma54; predicted +1 start at 3864692"
promoter complement(599. .626)
/note="factor Sigma70; promoter ibpB; documented+1 at 3865147"
promoter 647. .674
/note="factor Sigma70; predicted +1 start at 3865234"
gene 727. .1134
/gene="yidQ"
/note="b3688"
CDS 727. .1134
/gene="yidQ"
/function="orf; Unknown"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76711.1"
/db_xref="GI:1790123"
/translation="MMMLFSARDRYNSSGQHPRSTDEFIMIRNVLLAFMICSGMTLLG GCSSVMSHTGGKEGTYPGTRASATMIGDDETNWGTKSLAILDMPFTAVMDTLLLPWDV FRKDSSVRSRVEKSEANAQATNAVIPPARMPDN"
gene complement(1136. .2386)
/gene="yidR"
/note="b3689"
CDS complement(1136. .2386)
/gene="yidR"
/function="orf; Unknown"
/note="f416; 100 pct identical amino acid sequence and equal length to YIDR_ECOLI SW: P31455"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76712.1"
/db_xref="GI:1790124"
/translation="MMAGPVLYQDRAMKQITFAPRNHLLTNTNTWTPDSQWLVFDVRP SGASFTGETIERVNIHTGEVEVIYRASQGAHVGVVTVHPKSEKYVFIHGPENPDETWH YDFHHRRGVIVEGGKMSNLDAMDITAPYTPGVLRGGSHVHVFSPNGERVSFTYNDHVM HELDPALDLRNVGVAAPFGPVNVQKQHPREYSGSHWCVLVSKTTPTPQPGSDEINRAY EEGWVGNHALAFIGDTLSPKGEKVPELFIVELPQDEAGWKAAGDAPLSGTETTLPAPP RGVVQRRLTFTHHRAYPGLVNVPRHWVRCNPQGTQIAFLMRDDNGIVQLWLISPQGGE PRQLTHNKTDIQSAFNWHPSGEWLGFVLDNRIACAHAQSGEVEYLTEHHANSPSADAV VFSPDGQWLAWMEGGQLWITETDR"
promoter 2359. .2388
/note="factor Sigma70; predicted +1 start at 3866948"
gene 2430. .3515
/gene="yidS"
/note="b3690"
CDS 2430. .3515
/gene="yidS"
/function="orf; Unknown"
/note="o361; 100 pct identical amino acid sequence and

```
                     equal length to YIDS_ECOLI SW: P31456"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC76713.1"
                     /db_xref="GI:1790125"
                     /translation="MMNVSEKMEHFDVAIIGLGPAGSALARKLAGKMQVIALDKKHQC
                     GTEGFSKPCGGLLAPDAQRSFIRDGLTLPVDVIANPQIFSVKTVDVAASLTRNYQRSY
                     ININRHAFDLWMKSLIPASVEVYHDSLCRKIWREDDKWHVIFRADGWEQHITARYLVG
                     ADGANSMVRRHLYPDHQIRKYVAIQQWFAEKHPVPFYSCIFDNSITNCYSWSISKDGY
                     FIFGGAYPMKDGQTRFTTLKEKMSAFQFQFGKTVKSEKCTVLFPSRWQDFVCGKDNAF
                     LIGEAAGFISASSLEGISYALDSTDILRSVLLKQPEKLNTAYWRATRKLRLKLFGKIV
                     KSRCLTAPALRKWIMRSGVAHIPQLKD"
     promoter        complement(2436. .2464)
                     /note="factor Sigma70; predicted +1 start at 3866982"
     promoter        complement(2524. .2552)
                     /note="factor Sigma70; predicted +1 start at 3867070"
     gene            complement(3512. .4849)
                     /gene="dgoT"
                     /note="b3691"
     CDS             complement(3512. .4849)
                     /gene="dgoT"
                     /function="transport; Transport of small molecules:
                     Carbohydrates, organic acids, alcohols"
                     /note="f445; formerly designated yidT"
                     /codon_start=1
                     /transl_table=11
                     /product="D-galactonate transport"
                     /protein_id="AAC76714.1"
                     /db_xref="GI:1790126"
                     /translation="MVSGFAMPKIWRKLAMDIPVNAAKPGRRRYLTLVMIFITVVICY
                     VDRANLAVASAHIQEEFGITKAEMGYVFSAFAWLYTLCQIPGGWFLDRVGSRVTYFIA
                     IFGWSVATLFQGFATGLMSLIGLRAITGIFEAPAFPTNNRMVTSWFPEHERASAVGFY
                     TSGQFVGLAFLTPLLIWIQEMLSWHWVFIVTGGIGIIWSLIWFKVYQPPRLTKGISKA
                     ELDYIRDGGGLVDGDAPVKKEARQPLTAKDWKLVFHRKLIGVYLGQFAVASTLWFFLT
                     WFPNYLTQEKGITALKAGFMTTVPFLAAFVGVLLSGWVADLLVRKGFSLGFARKTPII
                     CGLLISTCIMGANYTNDPMMIMCLMALAFFGNGFASITWSLVSSLAPMRLIGLTGGVF
                     NFAGGLGGITVPLVVGYLAQGYGFAPALVYISAVALIGALSYILLVGDVKRVG"
     gene            complement(4924. .6687)
                     /gene="dgoA"
                     /note="b3692"
     CDS             complement(4924. .6687)
                     /gene="dgoA"
                     /function="enzyme; Degradation of small molecules: Carbon
                     compounds"
                     /note="f587; formerly designated yidU"
                     /codon_start=1
                     /transl_table=11
                     /product="2-oxo-3-deoxygalactonate 6-phosphate aldolase
                     and galactonate dehydratase"
                     /protein_id="AAC76715.1"
                     /db_xref="GI:2367262"
                     /translation="MQWQTKLPLIAILRGITPDEALAHVGAVIDAGFDAVEIPLNSPQ
                     WEQSIPAIVDAYGDKALIGAGTVLKPEQVDALARMGCQLIVTPNIHSEVIRRAVGYGM
                     TVCPGCATATEAFTALEAGAAGAENISVIGFWSAIHQSVKSGIAIGHRSLCRWRRDAE
                     NLAQWIDAGCAGAGLGSDLYRAGQSVERTAQQAAAFVKAYREAGAMKITKITTYRLPP

  Query Match           100.0%;  Score 64;  DB 1;  Length 10120;
  Best Local Similarity  100.0%;  Pred. No. 4e-11;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     650 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 591

Qy      61 AGCC 64
           ||||
Db     590 AGCC 587


RESULT 4
AE005600/c
LOCUS       AE005600               12642 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 219
            of 290.
ACCESSION   AE005600 AE005174
VERSION     AE005600.1  GI:12518521
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 12642)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 12642)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .12642
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            99. .1022
                     /gene="yidL"
                     /note="Z5175"
     CDS             99. .1022
                     /gene="yidL"
                     /function="putative regulator; Not classified"
                     /note="Residues 1 to 307 of 307 are 99.34 pct identical to
                     residues 1 to 307 of 307 from Escherichia coli K-12 Strain
                     MG1655: B3680"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ARAC-type regulatory protein"
                     /protein_id="AAG58883.1"
                     /db_xref="GI:12518522"
                     /translation="MSIIRRRDIKMNGKLQTSDVKNETPYNIPLLINENVISSGISLI
                     SLWHTYADEHYRVIWPRDKKKPLIANSWVAVYTVQGCGKILLKNGEQITLHGNCIIFL
                     KPMDIHSYHCEGLVWEQYWMEFTPTSMMDIPIGQQSVIYNGEIYNQELTEVAELITSP
                     EAIKNNLAVAFLTKIIYQWICLMYADGKKDPQRRQIEKLIATLHASLQQRWSVADMAA
                     TIPCSEAWLRRLFLRYTGKTPKEYYLDARLDLALSLLKQQGNSVGEVADTLNFFDSFH
                     FSKAFKHKFGYAPSAVLKNTDQHPTDASPHN"
     gene            complement(1019. .2341)
                     /gene="Z5177"
     CDS             complement(1019. .2341)
                     /gene="Z5177"
                     /function="putative enzyme; Transport of small molecules:
                     Carbohydrates, organic acids, alcohols"
                     /note="Residues 1 to 440 of 440 are 75.05 pct identical to
                     residues 1 to 441 of 441 from GenPept 118 :
                     gi|2145152|gb|AAB63015.1| (U81185) MalH [Fusobacterium
                     mortiferum]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative 6-phospho-beta-glucosidase"
                     /protein_id="AAG58884.1"
                     /db_xref="GI:12518523"
                     /translation="MTKFSVVVAGGGSTFTPGIVLMLLANQDRFPLRALKFYDNDGSR
```

```
                     QEVIAEACKVILKEKAPDIAFSYTTDPEVAFSDVDFVMAHIRVGKYPMRELDEKIPLR
                     HGVVGQETCGPGGIAYGMRSIGGVLELVDYMEKYSPNAWMLNYSNPAAIVAEATRRLR
                     PNAKILNICDMPIGIESRMAQIVGLQDRKQMRVRYYGLNHFGWWTSIEDLQGNDLMHQ
                     LRQYVSKHGYVPPQQDTHTEASWNDTYAKARDVQALDPDTLPNTYLKYYLFPDYVVQH
                     SNPEHTRANEVMEHREKQVFDACRAITAAGNSAAGKLEIDEHASYIVDLAAAIAFNTQ
                     ERMLLIVPNNGAIHNFDDEAMVEIPCLVGHNGPEPLVVGDIPQFQKGLMSQQVAVEKL
                     VVDAWEQRSYQHLWQAITLSKTVPSASVAKAILDELLEANKAYWPELR"
misc_feature         1241. .1748
                     /note="O-island #151; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655"
gene                 complement(2341. .3963)
                     /gene="Z5178"
CDS                  complement(2341. .3963)
                     /gene="Z5178"
                     /function="putative enzyme; Transport of small molecules:
                     Carbohydrates, organic acids, alcohols"
                     /note="Residues 1 to 370 of 540 are 99.45 pct identical to
                     residues 86 to 455 of 455 from Escherichia coli K-12
                     Strain MG1655: B3683"
                     /codon_start=1
                     /transl_table=11
                     /product="putative PTS component"
                     /protein_id="AAG58885.1"
                     /db_xref="GI:12518524"
                     /translation="MEMLSQIQRFGGAMFTPVLLFPFAGIVVGLAILLQNPMFVGESL
                     TDPNSLFAQIVHIIEEGGWTVFRNMPLIFAVGLPIGLAKQAQGRACLAVMVSFLTWNY
                     FINAMGMTWGSYFGVDFTQDAVAGSGLTMMAGIKTLDTSIIGAIIISGIVTALHNRLF
                     DKKLPVFLGIFQGTSYVVIIAFLVMIPCAWLTLLGWPKVQMGIESLQAFLRSAGALGV
                     WVYTFLERILIPTGLHHFIYGPFIFGPAAVEGGIQMYWAQHLQEFSLSAEPLKSLFPE
                     GGFALHGNSKIFGAVGISLAMYFTAAPENRVKVAGLLIPATLTAMLVGITEPLEFTFL
                     FISPLLFAVHAVLAASMSTVMYLFGVVGNMGGGLIDQVLPQNWIPMFSNHADMMLTQI
                     AIGLCFTLLYFVVFRTLILQFNMCTPGREDAEVKLYSKAEYKASRGQTTAAEPKKELD
                     QAAGILQALGGVGNISSINNCATRLRIALHDMSQTLDDEVFKKLGAHGVFRSGDAIQV
                     IIGLHVSQLREQLDSLINSHQSAENVAITEAV"
gene                 4253. .4969
                     /gene="yidP"
                     /note="Z5179"
CDS                  4253. .4969
                     /gene="yidP"
                     /function="putative regulator; Not classified"
                     /note="Residues 1 to 238 of 238 are 99.57 pct identical to
                     residues 1 to 238 of 238 from Escherichia coli K-12 Strain
                     MG1655: B3684"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /protein_id="AAG58886.1"
                     /db_xref="GI:12518525"
                     /translation="MIYKSIAERLRIRLNSADFTLNSLLPGEKKLAEEFAVSRMTIRK
                     AIDLLVAWGLVVRRHGSGTYLVRKDVLHQTASLTGLVEVLKRQGKTVTSQVLIFEIMP
                     APPAIASQLRIQINEQIYFSRRVRFVEGKPLMLEDSYMPVKLFRNLSLQHLEGSKFEY
                     IEQECGILIGGNYESLTPVLADRLLARQMKVAEHTPLLRITSLSYSESGEFLNYSVMF
                     RNASEYQVEYHLRRLHPDKS"
gene                 complement(4966. .6651)
                     /gene="yidE"
                     /note="Z5181"
CDS                  complement(4966. .6651)
                     /gene="yidE"
                     /function="putative transport; Not classified"
                     /note="Residues 1 to 561 of 561 are 100.00 pct identical
                     to residues 1 to 561 of 561 from Escherichia coli K-12
                     Strain MG1655: B3685"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transport protein"
                     /protein_id="AAG58887.1"
                     /db_xref="GI:12518526"
                     /translation="MLSQEKWTMSDIALTVSILALVAVVGLFIGNVKFRGIGLGIGGV
                     LFGGIIVGHFVSQAGMTLSSDMLHVIQEFGLILFVYTIGIQVGPGFFASLRVSGLRLN
                     LFAVLIVIIGGLVTAILHKLFDIPLPVVLGIFSGAVTNTPALGAGQQILRDLGTPMEM
                     VDQMGMSYAMAYPFGICGILFTMWMLRVIFRVNVETEAQQHESSRTNGGALIKTINIR
                     VENPNLHDLAIKDVPILNGDKIICSRLKREETLKVPSPDTIIQLGDLLHLVGQPADLH
                     NAQLVIGQEVDTSLSTKGTDLRVERVVVVTNENVLGKRIRDLHFKERYDVVISRLNRAG
                     VELVASGDISLQFGDILNLVGRPSAIDAVANVLGNAQQKLQQVQMLPVFIGIGLGVLL
                     GSIPVFVPGFPAALKLGLAGGPLIMALILGRIGSIGKLYWFMPPSANLALRELGIVLF
                     LSVVGLKSGGDFVNTLVNGEXLSWIGYGALITAVPLITVGILARMLAKMNYLTMCGML
                     AGSMTDPPALAFANNLHPTSGAAALSYATVYPLVMFLRIITPQLLAVLFWSIG"
gene                 complement(6823. .7257)
                     /gene="ibpB"
                     /note="Z5182"
CDS                  complement(6823. .7257)
                     /gene="ibpB"
                     /function="factor; Adaptations, atypical conditions"
                     /note="Residues 1 to 144 of 144 are 100.00 pct identical
                     to residues 1 to 144 of 144 from Escherichia coli K-12
                     Strain MG1655: B3686"
                     /codon_start=1
                     /transl_table=11
                     /product="heat shock protein"
                     /protein_id="AAG58888.1"
                     /db_xref="GI:12518527"
                     /translation="MTMRNFDLSPLMRQWIGFDKLANALQNAGESQSFPPYNIEKSDD
                     NHYRITLALAGFRQEDLEIQLEGTRLSVKGTPEQPKEEKKWLHQGLMNQPFSLSFTLA
                     ENMEVSGATFVNGLLHIDLIRNEPEPIAAQRIAISERPALNS"
gene                 complement(7363. .7776)
                     /gene="ibpA"
                     /note="Z5183"
CDS                  complement(7363. .7776)
                     /gene="ibpA"
                     /function="factor; Adaptations, atypical conditions"
                     /note="Residues 1 to 137 of 137 are 100.00 pct identical
                     to residues 1 to 137 of 137 from Escherichia coli K-12
                     Strain MG1655: B3687"
                     /codon_start=1
                     /transl_table=11
                     /product="heat shock protein"
                     /protein_id="AAG58889.1"
                     /db_xref="GI:12518528"
                     /translation="MRNFDLSPLYRSAIGFDRLFNHLENNQSQSNGGYPPYNVELVDE
                     NHYRIAIAVAGFAESELEITAQDNLLVVKGAHADEQKERTYLYQGIAERNFERKFQLA
                     ENIHVRGANLVNGLLYIDLERVIPEAKKPRRIEIN"
gene                 8007. .8414
                     /gene="yidQ"
                     /note="Z5184"
CDS                  8007. .8414
                     /gene="yidQ"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 135 of 135 are 99.25 pct identical to
                     residues 1 to 135 of 135 from Escherichia coli K-12 Strain
                     MG1655: B3688"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG58890.1"
                     /db_xref="GI:12518529"
                     /translation="MMMLFSARDRYNSSGQHPRSTDEFIMIRNVLLAFMICSGMALLG
                     GCSSVMSHTGGKEGTYPGTRASATMIGDDETNWGTKSLAILDMPFTAVMDTLLLPWDV

 Query Match          100.0%;  Score 64;  DB 1;  Length 12642;
 Best Local Similarity  100.0%;  Pred. No. 3.9e-11;
 Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7930 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 7871

Qy      61 AGCC 64
           ||||
Db    7870 AGCC 7867


RESULT 5
ECOUW82/c
LOCUS       ECOUW82               136254 bp    DNA     linear   BCT 07-FEB-1995
DEFINITION  E. coli; the region from 81.5 to 84.5 minutes.
ACCESSION   L10328
VERSION     L10328.1  GI:290484
```

KEYWORDS .
SOURCE Escherichia coli K12 strain MG1655; lambda clones EC14-52, EC17-187, EC27-890, EC21-99, EC22-175, EC27-236, EC17-30, EC17-137, EC19-71, EC15-15, EC17-159; subclones in M13mp19 or Janus.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 136254)
AUTHORS Burland,V., Plunkett,G. III, Daniels,D.L. and Blattner,F.R.
TITLE DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication
JOURNAL Genomics 16 (3), 551-561 (1993)
MEDLINE 93315143
PUBMED 7686882
COMMENT This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. A preliminary report was presented at the Hilton Head meeting'Genome Sequencing and Analysis IV' September 26 - 30,1992. The entire sequence was independently determined from E. coli MG1655; overlaps and conflicts with other sequence determinations are annotated. The end of this entry overlaps the start of the entry ECOUW85 (M87049) by the six bases of an EcoRI site.
Data kindly submitted in computer readable form by: Guy Plunkett III
Laboratory of Genetics
University of Wisconsin
445 Henry Mall
Madison WI 53706
USA
Phone: 608-262-2534
Email: ecoli@genetics.wisc.edu
Fax: 608-263-7459
[1] Authors request hold until publication.

FEATURES Location/Qualifiers

source 1. .136254
/organism="Escherichia coli"
/db_xref="taxon:562"

misc_feature 1. .11607
/note="EC14-52 (lambda clone)"

misc_feature complement(<1. .669)
/note="X06036; ECOFPG(1. .>669)"

gene complement(1. .544)
/gene="fpg"

CDS complement(<1. .544)
/gene="fpg"
/note="alternate gene name mutH"
/codon_start=1
/transl_table=11
/product="formamidopyrimidine-DNA glycosylase"
/protein_id="AAA61988.1"
/db_xref="GI:643609"
/translation="MPELPEVETSRRGIEPHLVGATILHAVVRNGRLRWPVSEEIYRL
SDQPVLSVQRRAKYLLLELPEGWIIIHLGMSGSLRILPEELPPEKHDHVDLVMSNGKV
LRYTDPRRFGAWLWTKELEGHNVLTHLGPEPLSDDFNGEYLHQKCAKKKTAIKPWLMD
NKLVVGVGNIYASESLFAAGI"

misc_feature complement(526. .1290)
/note="J01677; ECORPMBG(1. .764)"

promoter complement(557. .586)
/note="promoter-like sequence; promoter matrix score of 44"

misc_difference 596. .601
/note="AAAAA in X06036; AAAAAA in J01677 and here"

misc_difference 611
/note="G in J01677; A in X06036 and here"

misc_difference 615
/note="G in J01677; A in X06036 and here"

misc_difference 618. .623
/note="TTTTTTT in X06036; TTTTTT in J01677 and here"

gene complement(642. .809)
/gene="rpmG (CG Site No. 236)"

CDS complement(642. .809)
/gene="rpmG (CG Site No. 236)"
/codon_start=1
/transl_table=11
/product="50S ribosomal subunit protein L33"
/protein_id="AAA61989.1"
/db_xref="GI:290486"
/translation="MAKGIREKIKLVSSAGTGHFYTTTKNKRTKPEKLELKKFDPVVR
QHVIYKEAKIK"

gene complement(830. .1066)
/gene="rpmB (CG Site No. 240)"

CDS complement(830. .1066)
/gene="rpmB (CG Site No. 240)"
/codon_start=1
/transl_table=11
/product="50S ribosomal subunit protein L28"
/protein_id="AAA61990.1"
/db_xref="GI:290487"
/translation="MSRVCQVTGKRPVTGNNRSHALNATKRRFLPNLHSHRFWVESEK
RFVTLRVSAKGMRVIDKKGIDTVLAELRARGEKY"

misc_difference 902
/note="G in J01677; A here"

misc_difference 1128
/note="TC in J01677; C here"

misc_difference 1153. .1155
/note="T in J01677; TAT here"

misc_feature complement(1168. .1819)
/note="X63366; ECORADC(1. .651)"

promoter complement(1212. .1240)
/note="promoter-like sequence; promoter matrix score of 77"

gene complement(1283. .1957)
/gene="radC (CG Site No. 13913)"

CDS complement(1283. .1957)
/gene="radC (CG Site No. 13913)"
/note="differs from radC sequence in X63366; similar to S. aureus putative 25 kD protein"
/codon_start=1
/transl_table=11
/protein_id="AAA61991.1"
/db_xref="GI:290488"
/translation="MKVKNNSQLLMPREKMLKFGISALTDVELLALFLRTGTRGKDVL
TLAKEMLENFGSLYGLLTSEYEQFSGVHGIGVAKFAQLKGIAELARRYYNVRMREESP
LLSPEMTREFLQSQLTGEEREIFMVIFLDSQHRVITHRRLFSGTLNHVEVHPREIIRE
AIKINASALILAHNHPSGCAEPSKADKLITERIIKSCQFMDLRVLDHIVIGRGEYVSF
AERGWI"

misc_difference 1369. .1370
/note="A in X63366; AA here"

variation 1436
/note="A deleted in radC102 mutation (X63367)"

misc_structure 1698. .1822
/note="predicted bend of 76 degrees"

promoter 1875. .1902
/note="promoter-like sequence; promoter matrix score of 50"

promoter complement(1911. .1940)
/gene="radC (CG Site No. 13913)"
/note="promoter-like sequence; promoter matrix score of 45"

gene 2051. .3343
/gene="dfp (CG Site No. 18430)"

CDS 2051. .3343
/gene="dfp (CG Site No. 18430)"
/codon_start=1
/transl_table=11
/protein_id="AAA61992.1"
/db_xref="GI:290489"
/translation="MKARQQKYCDKIANFWCHPTGKIIMSLAGKKIVLGVSGGIAAYK
TPELVRRLRDRGADVRVAMTEAAKAFITPLSLQAVSGYPVSDSLLDPAAEAAMGHIEL
GKWADLVILAPATADLIARVAAGMANDLVSTICLATPAPVAVLPAMNQQMYRAAATQH
NLEVLASRGLLIWGPDSGSQACGDIGPGRXXDPLTIVDMAVAHFSPVNDLKHLNIMIT
AGPTREPLDPVRYISNHSSGKMGFAIAAAAARRGANVTLVSGPVSLPTPPFVKRVDVM
TALEMEAAVNASVQQQNIFIGCAAVADYRAATVAPEKIKKQATQGDELTIKMVKNPDI
VAGVAALKDHRPYVVGFAAETNNVEEYARQKRIRKNLDLICANDVSQPTQGFNSDNNA

```
                     LHLFWQDGDKVLPLERKELLGQLLLDEIVTRYDEKNRR"
     misc_feature    2982. .5552
                     /note="V01578; ECODUTPYR(1. .2568)"
     misc_feature    2982. .4591
                     /note="X01714; ECOPYREA(1. .1609)"
     gene            3324. .3779
                     /gene="dut (CG Site No. 832)"
     CDS             3324. .3779
                     /gene="dut (CG Site No. 832)"
                     /codon_start=1
                     /transl_table=11
                     /product="deoxyuridinetriphosphatase"
                     /protein_id="AAA61993.1"
                     /db_xref="GI:290490"
                     /translation="MKKIDVKILDPRVGKEFPLPTYATSGSAGLDLRACLNDAVELAP
                     GDTTLVPTGLAIHIADPSLAAMMLPRSGLGHKHGIVLGNLVGLIDSDYQGQLMISVWN
                     RGQDSFTIQPGERIAQMIFVPVVQAEFNLVEDFDATDRGEGGFGHSGRQ"
     gene            3844. .4482
                     /gene="ttk"
     CDS             3844. .4482
                     /gene="ttk"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="23 kD protein"
                     /protein_id="AAA61994.1"
                     /db_xref="GI:290491"
                     /translation="MPPGKCLFSGVFCNMAEKQTAKRNRREEILQSLALMLESSDGSQ
                     RITTAKLAASVGVSEAALYRHFPSKTRMFDSLIEFIEDSLITRINLILKDEKDTTARL
                     RLIVLLLLGFGERNPGLTRILTGHALMFEQDRLQGRINQLFERIEAQLRQVLREKRMR
                     EGEGYTTDETLLASQILAFCEGMLSRFVRSEFKYRPTDDFDARWPLIAAQLQ"
     misc_difference 4469. .4471
                     /gene="ttk"
                     /note="CC in X01714 and V01578; CAC here"
     repeat_region   4489. .4514
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 1 REP sequence"
     gene            complement(4519. .5160)
                     /gene="pyrE (CG Site No. 327)"
     CDS             complement(4519. .5160)
                     /gene="pyrE (CG Site No. 327)"
                     /codon_start=1
                     /transl_table=11
                     /product="orotate phosphoribosyltransferase"
                     /protein_id="AAA61995.1"
                     /db_xref="GI:643610"
                     /translation="MKPYQRQFIEFALSKQVLKFGEFTLKSGRKSPYFFNAGLFNTGR
                     DLALLGRFYAEALVDSGIEFDLLFGPAYKGIPIATTTAVALAEHHDLDLPYCFNRKEA
                     KDHGEGGNLVGSALQGRVMLVDDVITAGTAIRESMEIIQANGATLAGVLISLDRQERG
                     RGEISAIQEVERDYNCKVISIITLKDLIAYLEEKPEMAEHLAAVKAYREEFGV"

  Query Match          100.0%;  Score 64;  DB 1;  Length 136254;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-11;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56970 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 56911

Qy      61 AGCC 64
           ||||
Db   56910 AGCC 56907


RESULT 6
AC023611
LOCUS       AC023611               221285 bp    DNA     linear   HTG 29-JUN-2000
DEFINITION  Mus musculus clone CT7-9K21, WORKING DRAFT SEQUENCE, 62 unordered
            pieces.
ACCESSION   AC023611
VERSION     AC023611.2  GI:8810277
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 221285)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221285)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jun 29, 2000 this sequence version replaced gi:6980221.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 956898
            Center clone name: RG-MBAC_9K21
            -------
            Summary Statistics
            Consensus quality: 172196 bases at least Q40
            Consensus quality: 196011 bases at least Q30
            Consensus quality: 201702 bases at least Q20
            Estimated insert size: 160000; pulse field gel estimation
            Estimated insert size: 215185; sum-of-contigs estimation
            Quality coverage:  6.26 in Q20 bases; pulse field gel estimation
            Quality coverage:  4.65 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 62 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1033: contig of 1033 bp in length
            *     1034     1133: gap of unknown length
            *     1134     2542: contig of 1409 bp in length
            *     2543     2642: gap of unknown length
            *     2643     3698: contig of 1056 bp in length
            *     3699     3798: gap of unknown length
            *     3799     4993: contig of 1195 bp in length
            *     4994     5093: gap of unknown length
            *     5094     6427: contig of 1334 bp in length
            *     6428     6527: gap of unknown length
            *     6528     7763: contig of 1236 bp in length
            *     7764     7863: gap of unknown length
            *     7864     8877: contig of 1014 bp in length
            *     8878     8977: gap of unknown length
            *     8978    10256: contig of 1279 bp in length
            *    10257    10356: gap of unknown length
            *    10357    11356: contig of 1000 bp in length
            *    11357    11456: gap of unknown length
            *    11457    12470: contig of 1014 bp in length
            *    12471    12570: gap of unknown length
            *    12571    13861: contig of 1291 bp in length
            *    13862    13961: gap of unknown length
            *    13962    15060: contig of 1099 bp in length
            *    15061    15160: gap of unknown length
            *    15161    16252: contig of 1092 bp in length
            *    16253    16352: gap of unknown length
            *    16353    17470: contig of 1118 bp in length
            *    17471    17570: gap of unknown length
            *    17571    18837: contig of 1267 bp in length
            *    18838    18937: gap of unknown length
            *    18938    20246: contig of 1309 bp in length
            *    20247    20346: gap of unknown length
            *    20347    21515: contig of 1169 bp in length
            *    21516    21615: gap of unknown length
            *    21616    22752: contig of 1137 bp in length
```

```
*    22753    22852: gap of unknown length
*    22853    24438: contig of 1586 bp in length
*    24439    24538: gap of unknown length
*    24539    25615: contig of 1077 bp in length
*    25616    25715: gap of unknown length
*    25716    26918: contig of 1203 bp in length
*    26919    27018: gap of unknown length
*    27019    28161: contig of 1143 bp in length
*    28162    28261: gap of unknown length
*    28262    29906: contig of 1645 bp in length
*    29907    30006: gap of unknown length
*    30007    31335: contig of 1329 bp in length
*    31336    31435: gap of unknown length
*    31436    32661: contig of 1226 bp in length
*    32662    32761: gap of unknown length
*    32762    34093: contig of 1332 bp in length
*    34094    34193: gap of unknown length
*    34194    35482: contig of 1289 bp in length
*    35483    35582: gap of unknown length
*    35583    37001: contig of 1419 bp in length
*    37002    37101: gap of unknown length
*    37102    38337: contig of 1236 bp in length
*    38338    38437: gap of unknown length
*    38438    39825: contig of 1388 bp in length
*    39826    39925: gap of unknown length
*    39926    41162: contig of 1237 bp in length
*    41163    41262: gap of unknown length
*    41263    42872: contig of 1610 bp in length
*    42873    42972: gap of unknown length
*    42973    44118: contig of 1146 bp in length
*    44119    44218: gap of unknown length
*    44219    45913: contig of 1695 bp in length
*    45914    46013: gap of unknown length
*    46014    47295: contig of 1282 bp in length
*    47296    47395: gap of unknown length
*    47396    48500: contig of 1105 bp in length
*    48501    48600: gap of unknown length
*    48601    50670: contig of 2070 bp in length
*    50671    50770: gap of unknown length
*    50771    52744: contig of 1974 bp in length
*    52745    52844: gap of unknown length
*    52845    54280: contig of 1436 bp in length
*    54281    54380: gap of unknown length
*    54381    56311: contig of 1931 bp in length
*    56312    56411: gap of unknown length
*    56412    59166: contig of 2755 bp in length
*    59167    59266: gap of unknown length
*    59267    62681: contig of 3415 bp in length
*    62682    62781: gap of unknown length
*    62782    65691: contig of 2910 bp in length
*    65692    65791: gap of unknown length
*    65792    69358: contig of 3567 bp in length
*    69359    69458: gap of unknown length
*    69459    71957: contig of 2499 bp in length
*    71958    72057: gap of unknown length
*    72058    76352: contig of 4295 bp in length
*    76353    76452: gap of unknown length
*    76453    79737: contig of 3285 bp in length
*    79738    79837: gap of unknown length
*    79838    86357: contig of 6520 bp in length
*    86358    86457: gap of unknown length
*    86458    90914: contig of 4457 bp in length
*    90915    91014: gap of unknown length
*    91015    96832: contig of 5818 bp in length
*    96833    96932: gap of unknown length
*    96933   103782: contig of 6850 bp in length
*   103783   103882: gap of unknown length
*   103883   110628: contig of 6746 bp in length
*   110629   110728: gap of unknown length
*   110729   116501: contig of 5773 bp in length
*   116502   116601: gap of unknown length
*   116602   123034: contig of 6433 bp in length
*   123035   123134: gap of unknown length
*   123135   130209: contig of 7075 bp in length
*   130210   130309: gap of unknown length
*   130310   141462: contig of 11153 bp in length
*   141463   141562: gap of unknown length
*   141563   149389: contig of 7827 bp in length
*   149390   149489: gap of unknown length
*   149490   157306: contig of 7817 bp in length
*   157307   157406: gap of unknown length
*   157407   167404: contig of 9998 bp in length
*   167405   167504: gap of unknown length
*   167505   185812: contig of 18308 bp in length
*   185813   185912: gap of unknown length
*   185913   205026: contig of 19114 bp in length
*   205027   205126: gap of unknown length
*   205127   221285: contig of 16159 bp in length.
FEATURES             Location/Qualifiers
     source          1. .221285
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="CT7-9K21"
                     /clone_lib="CitbCJ7 mouse BAC library"
BASE COUNT    61589 a  46549 c  45795 g  61182 t   6170 others
ORIGIN

  Query Match           100.0%;  Score 64;  DB 2;  Length 221285;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-11;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42313 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 42372

Qy      61 AGCC 64
          ||||
Db   42373 AGCC 42376


RESULT 7
AC073670/c
LOCUS       AC073670               252977 bp    DNA     linear   HTG 08-AUG-2000
DEFINITION  Mus musculus clone CT7-368A6, WORKING DRAFT SEQUENCE, 85 unordered
            pieces.
ACCESSION   AC073670
VERSION     AC073670.1  GI:8810287
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 252977)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252977)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1094499
            Center clone name: RG-MBAC_368A6
            -------
            Summary Statistics
            Consensus quality: 177673 bases at least Q40
            Consensus quality: 214946 bases at least Q30
            Consensus quality: 224156 bases at least Q20
            Estimated insert size: 100000; pulse field gel estimation
            Estimated insert size: 244577; sum-of-contigs estimation
```

Quality coverage: 6.6 in Q20 bases; pulse field gel estimation
Quality coverage: 2.7 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 85 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1110: contig of 1110 bp in length
*	1111	1210: gap of unknown length
*	1211	2457: contig of 1247 bp in length
*	2458	2557: gap of unknown length
*	2558	3663: contig of 1106 bp in length
*	3664	3763: gap of unknown length
*	3764	4896: contig of 1133 bp in length
*	4897	4996: gap of unknown length
*	4997	6152: contig of 1156 bp in length
*	6153	6252: gap of unknown length
*	6253	7288: contig of 1036 bp in length
*	7289	7388: gap of unknown length
*	7389	8964: contig of 1576 bp in length
*	8965	9064: gap of unknown length
*	9065	10174: contig of 1110 bp in length
*	10175	10274: gap of unknown length
*	10275	11586: contig of 1312 bp in length
*	11587	11686: gap of unknown length
*	11687	12873: contig of 1187 bp in length
*	12874	12973: gap of unknown length
*	12974	14204: contig of 1231 bp in length
*	14205	14304: gap of unknown length
*	14305	15841: contig of 1537 bp in length
*	15842	15941: gap of unknown length
*	15942	17159: contig of 1218 bp in length
*	17160	17259: gap of unknown length
*	17260	18555: contig of 1296 bp in length
*	18556	18655: gap of unknown length
*	18656	19742: contig of 1087 bp in length
*	19743	19842: gap of unknown length
*	19843	21273: contig of 1431 bp in length
*	21274	21373: gap of unknown length
*	21374	22750: contig of 1377 bp in length
*	22751	22850: gap of unknown length
*	22851	24000: contig of 1150 bp in length
*	24001	24100: gap of unknown length
*	24101	25538: contig of 1438 bp in length
*	25539	25638: gap of unknown length
*	25639	26726: contig of 1088 bp in length
*	26727	26826: gap of unknown length
*	26827	28011: contig of 1185 bp in length
*	28012	28111: gap of unknown length
*	28112	29531: contig of 1420 bp in length
*	29532	29631: gap of unknown length
*	29632	31152: contig of 1521 bp in length
*	31153	31252: gap of unknown length
*	31253	32452: contig of 1200 bp in length
*	32453	32552: gap of unknown length
*	32553	33654: contig of 1102 bp in length
*	33655	33754: gap of unknown length
*	33755	34774: contig of 1020 bp in length
*	34775	34874: gap of unknown length
*	34875	36051: contig of 1177 bp in length
*	36052	36151: gap of unknown length
*	36152	37590: contig of 1439 bp in length
*	37591	37690: gap of unknown length
*	37691	38908: contig of 1218 bp in length
*	38909	39008: gap of unknown length
*	39009	40601: contig of 1593 bp in length
*	40602	40701: gap of unknown length
*	40702	41926: contig of 1225 bp in length
*	41927	42026: gap of unknown length
*	42027	43983: contig of 1957 bp in length
*	43984	44083: gap of unknown length
*	44084	45206: contig of 1123 bp in length
*	45207	45306: gap of unknown length
*	45307	46881: contig of 1575 bp in length
*	46882	46981: gap of unknown length
*	46982	48056: contig of 1075 bp in length
*	48057	48156: gap of unknown length
*	48157	49668: contig of 1512 bp in length
*	49669	49768: gap of unknown length
*	49769	51187: contig of 1419 bp in length
*	51188	51287: gap of unknown length
*	51288	52332: contig of 1045 bp in length
*	52333	52432: gap of unknown length
*	52433	53754: contig of 1322 bp in length
*	53755	53854: gap of unknown length
*	53855	55581: contig of 1727 bp in length
*	55582	55681: gap of unknown length
*	55682	57262: contig of 1581 bp in length
*	57263	57362: gap of unknown length
*	57363	59152: contig of 1790 bp in length
*	59153	59252: gap of unknown length
*	59253	60732: contig of 1480 bp in length
*	60733	60832: gap of unknown length
*	60833	63030: contig of 2198 bp in length
*	63031	63130: gap of unknown length
*	63131	65366: contig of 2236 bp in length
*	65367	65466: gap of unknown length
*	65467	67227: contig of 1761 bp in length
*	67228	67327: gap of unknown length
*	67328	68506: contig of 1179 bp in length
*	68507	68606: gap of unknown length
*	68607	71259: contig of 2653 bp in length
*	71260	71359: gap of unknown length
*	71360	73965: contig of 2606 bp in length
*	73966	74065: gap of unknown length
*	74066	76125: contig of 2060 bp in length
*	76126	76225: gap of unknown length
*	76226	77297: contig of 1072 bp in length
*	77298	77397: gap of unknown length
*	77398	79646: contig of 2249 bp in length
*	79647	79746: gap of unknown length
*	79747	82449: contig of 2703 bp in length
*	82450	82549: gap of unknown length
*	82550	85345: contig of 2796 bp in length
*	85346	85445: gap of unknown length
*	85446	86817: contig of 1372 bp in length
*	86818	86917: gap of unknown length
*	86918	90145: contig of 3228 bp in length
*	90146	90245: gap of unknown length
*	90246	91490: contig of 1245 bp in length
*	91491	91590: gap of unknown length
*	91591	94935: contig of 3345 bp in length
*	94936	95035: gap of unknown length
*	95036	97845: contig of 2810 bp in length
*	97846	97945: gap of unknown length
*	97946	100449: contig of 2504 bp in length
*	100450	100549: gap of unknown length
*	100550	103583: contig of 3034 bp in length
*	103584	103683: gap of unknown length
*	103684	107678: contig of 3995 bp in length
*	107679	107778: gap of unknown length
*	107779	110534: contig of 2756 bp in length
*	110535	110634: gap of unknown length
*	110635	113491: contig of 2857 bp in length
*	113492	113591: gap of unknown length
*	113592	115603: contig of 2012 bp in length
*	115604	115703: gap of unknown length
*	115704	118957: contig of 3254 bp in length
*	118958	119057: gap of unknown length
*	119058	124203: contig of 5146 bp in length
*	124204	124303: gap of unknown length
*	124304	127811: contig of 3508 bp in length
*	127812	127911: gap of unknown length

```
*   127912   132157: contig of 4246 bp in length
*   132158   132257: gap of unknown length
*   132258   135753: contig of 3496 bp in length
*   135754   135853: gap of unknown length
*   135854   140373: contig of 4520 bp in length
*   140374   140473: gap of unknown length
*   140474   145161: contig of 4688 bp in length
*   145162   145261: gap of unknown length
*   145262   149776: contig of 4515 bp in length
*   149777   149876: gap of unknown length
*   149877   156082: contig of 6206 bp in length
*   156083   156182: gap of unknown length
*   156183   163129: contig of 6947 bp in length
*   163130   163229: gap of unknown length
*   163230   167273: contig of 4044 bp in length
*   167274   167373: gap of unknown length
*   167374   175126: contig of 7753 bp in length
*   175127   175226: gap of unknown length
*   175227   182114: contig of 6888 bp in length
*   182115   182214: gap of unknown length

  Query Match             100.0%;  Score 64;  DB 2;  Length 252977;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-11;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11123 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 11064

Qy      61 AGCC 64
           ||||
Db   11063 AGCC 11060


RESULT 8
AC020870/c
LOCUS       AC020870               256373 bp    DNA     linear   HTG 17-FEB-2000
DEFINITION  Mus musculus clone RP23-302J15, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC020870
VERSION     AC020870.2  GI:6984372
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 256373)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 256373)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Feb 17, 2000 this sequence version replaced gi:6686438.
            * NOTE: This record contains 198 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      752: contig of 752 bp in length
            *                    gap of unknown length
            *      753     1000: contig of 248 bp in length
            *                    gap of unknown length
            *     1001     1604: contig of 604 bp in length
            *                    gap of unknown length
            *     1605     2302: contig of 698 bp in length
            *                    gap of unknown length
            *     2303     2699: contig of 397 bp in length
            *                    gap of unknown length
            *     2700     3686: contig of 987 bp in length
            *                    gap of unknown length
            *     3687     4099: contig of 413 bp in length
            *                    gap of unknown length
            *     4100     4778: contig of 679 bp in length
            *                    gap of unknown length
            *     4779     5571: contig of 793 bp in length
            *                    gap of unknown length
            *     5572     5967: contig of 396 bp in length
            *                    gap of unknown length
            *     5968     6763: contig of 796 bp in length
            *                    gap of unknown length
            *     6764     7639: contig of 876 bp in length
            *                    gap of unknown length
            *     7640     8442: contig of 803 bp in length
            *                    gap of unknown length
            *     8443     9414: contig of 972 bp in length
            *                    gap of unknown length
            *     9415    10236: contig of 822 bp in length
            *                    gap of unknown length
            *    10237    10844: contig of 608 bp in length
            *                    gap of unknown length
            *    10845    11504: contig of 660 bp in length
            *                    gap of unknown length
            *    11505    12321: contig of 817 bp in length
            *                    gap of unknown length
            *    12322    12652: contig of 331 bp in length
            *                    gap of unknown length
            *    12653    14116: contig of 1464 bp in length
            *                    gap of unknown length
            *    14117    15298: contig of 1182 bp in length
            *                    gap of unknown length
            *    15299    16234: contig of 936 bp in length
            *                    gap of unknown length
            *    16235    16376: contig of 142 bp in length
            *                    gap of unknown length
            *    16377    16664: contig of 288 bp in length
            *                    gap of unknown length
            *    16665    17631: contig of 967 bp in length
            *                    gap of unknown length
            *    17632    19241: contig of 1610 bp in length
            *                    gap of unknown length
            *    19242    20453: contig of 1212 bp in length
            *                    gap of unknown length
            *    20454    20559: contig of 106 bp in length
            *                    gap of unknown length
            *    20560    21117: contig of 558 bp in length
            *                    gap of unknown length
            *    21118    21655: contig of 538 bp in length
            *                    gap of unknown length
            *    21656    21818: contig of 163 bp in length
            *                    gap of unknown length
            *    21819    22679: contig of 861 bp in length
            *                    gap of unknown length
            *    22680    23579: contig of 900 bp in length
            *                    gap of unknown length
            *    23580    24383: contig of 804 bp in length
            *                    gap of unknown length
            *    24384    25612: contig of 1229 bp in length
            *                    gap of unknown length
            *    25613    25957: contig of 345 bp in length
            *                    gap of unknown length
            *    25958    27260: contig of 1303 bp in length
            *                    gap of unknown length
            *    27261    28596: contig of 1336 bp in length
            *                    gap of unknown length
            *    28597    29473: contig of 877 bp in length
            *                    gap of unknown length
            *    29474    30372: contig of 899 bp in length
            *                    gap of unknown length
```

```
*    30373    31202: contig of 830 bp in length
*                 gap of unknown length
*    31203    32688: contig of 1486 bp in length
*                 gap of unknown length
*    32689    33431: contig of 743 bp in length
*                 gap of unknown length
*    33432    33693: contig of 262 bp in length
*                 gap of unknown length
*    33694    34616: contig of 923 bp in length
*                 gap of unknown length
*    34617    35884: contig of 1268 bp in length
*                 gap of unknown length
*    35885    36459: contig of 575 bp in length
*                 gap of unknown length
*    36460    36996: contig of 537 bp in length
*                 gap of unknown length
*    36997    37941: contig of 945 bp in length
*                 gap of unknown length
*    37942    38682: contig of 741 bp in length
*                 gap of unknown length
*    38683    39549: contig of 867 bp in length
*                 gap of unknown length
*    39550    40792: contig of 1243 bp in length
*                 gap of unknown length
*    40793    42585: contig of 1793 bp in length
*                 gap of unknown length
*    42586    44115: contig of 1530 bp in length
*                 gap of unknown length
*    44116    45300: contig of 1185 bp in length
*                 gap of unknown length
*    45301    46417: contig of 1117 bp in length
*                 gap of unknown length
*    46418    47815: contig of 1398 bp in length
*                 gap of unknown length
*    47816    49310: contig of 1495 bp in length
*                 gap of unknown length
*    49311    50304: contig of 994 bp in length
*                 gap of unknown length
*    50305    50848: contig of 544 bp in length
*                 gap of unknown length
*    50849    51828: contig of 980 bp in length
*                 gap of unknown length
*    51829    52699: contig of 871 bp in length
*                 gap of unknown length
*    52700    53365: contig of 666 bp in length
*                 gap of unknown length
*    53366    54633: contig of 1268 bp in length
*                 gap of unknown length
*    54634    55470: contig of 837 bp in length
*                 gap of unknown length
*    55471    56358: contig of 888 bp in length
*                 gap of unknown length
*    56359    56934: contig of 576 bp in length
*                 gap of unknown length
*    56935    57672: contig of 738 bp in length
*                 gap of unknown length
*    57673    58319: contig of 647 bp in length
*                 gap of unknown length
*    58320    58529: contig of 210 bp in length
*                 gap of unknown length
*    58530    59083: contig of 554 bp in length
*                 gap of unknown length
*    59084    59906: contig of 823 bp in length
*                 gap of unknown length
*    59907    60826: contig of 920 bp in length
*                 gap of unknown length
*    60827    61084: contig of 258 bp in length
*                 gap of unknown length
*    61085    61762: contig of 678 bp in length
*                 gap of unknown length
*    61763    62638: contig of 876 bp in length
*                 gap of unknown length
*    62639    62713: contig of 75 bp in length
*                 gap of unknown length
*    62714    63593: contig of 880 bp in length
*                 gap of unknown length
*    63594    64039: contig of 446 bp in length
*                 gap of unknown length
*    64040    65076: contig of 1037 bp in length
*                 gap of unknown length
*    65077    65366: contig of 290 bp in length
*                 gap of unknown length
*    65367    65590: contig of 224 bp in length
*                 gap of unknown length
*    65591    66041: contig of 451 bp in length
*                 gap of unknown length
*    66042    67040: contig of 999 bp in length
*                 gap of unknown length
*    67041    67560: contig of 520 bp in length
*                 gap of unknown length

 Query Match           100.0%;  Score 64;  DB 2;  Length 256373;
 Best Local Similarity  100.0%;  Pred. No. 2.3e-11;
 Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31717 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 31658

Qy      61 AGCC 64
           ||||
Db   31657 AGCC 31654


RESULT 9
AC090533
LOCUS       AC090533               272545 bp    DNA     linear   HTG 15-MAY-2002
DEFINITION  Mus musculus clone RP23-33H13 strain C57BL6/J, WORKING DRAFT
            SEQUENCE, 45 unordered pieces.
ACCESSION   AC090533
VERSION     AC090533.4  GI:17976439
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 272545)
  AUTHORS   Li,L., Montgomery,K.T., Grills,G., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Han,J., Ioshikhes,I.P., Lee,E.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 272545)
  AUTHORS   Li,L., Montgomery,K.T., Grills,G., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Han,J., Ioshikhes,I.P., Lee,E.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAR-2001) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10461, USA
COMMENT     On Dec 21, 2001 this sequence version replaced gi:13310871.
            ----------------Genome Center
            Center:  Harvard Partners Genome Center
            Center Code:  HPGC
            Web site: http://www.hpcgg.org/Sequence/mouse.html
            Contact: hpgc@mendel.mgh.harvard.edu
            ----------------Summary Statistics
            Center project name:  ADU
            Sequencing vector:  pUC18; L08752
            Chemistry:  Dye-terminator Big Dye; 100%
            *Consensus quality:  257441 at least Q20
            *Consensus quality:  253718 at least Q30
            *Consensus quality:  247095 at least Q40
            Estimated insert size:  agarose-FP - N/A
            **Estimated insert size: 271665 - sum-of-contigs
            Quality coverage:  agarose-FP - N/A
```

Quality coverage: 6.8 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	30267: contig of 30267 bp in length
*	30268	30287: gap of unknown length
*	30288	58400: contig of 28113 bp in length
*	58401	58420: gap of unknown length
*	58421	88864: contig of 30444 bp in length
*	88865	88884: gap of unknown length
*	88885	109889: contig of 21005 bp in length
*	109890	109909: gap of unknown length
*	109910	125133: contig of 15224 bp in length
*	125134	125153: gap of unknown length
*	125154	139549: contig of 14396 bp in length
*	139550	139569: gap of unknown length
*	139570	153410: contig of 13841 bp in length
*	153411	153430: gap of unknown length
*	153431	162567: contig of 9137 bp in length
*	162568	162587: gap of unknown length
*	162588	171044: contig of 8457 bp in length
*	171045	171064: gap of unknown length
*	171065	179304: contig of 8240 bp in length
*	179305	179324: gap of unknown length
*	179325	187059: contig of 7735 bp in length
*	187060	187079: gap of unknown length
*	187080	197598: contig of 10519 bp in length
*	197599	197618: gap of unknown length
*	197619	205966: contig of 8348 bp in length
*	205967	205986: gap of unknown length
*	205987	212449: contig of 6463 bp in length
*	212450	212469: gap of unknown length
*	212470	217752: contig of 5283 bp in length
*	217753	217772: gap of unknown length
*	217773	223943: contig of 6171 bp in length
*	223944	223963: gap of unknown length
*	223964	230362: contig of 6399 bp in length
*	230363	230382: gap of unknown length
*	230383	236251: contig of 5869 bp in length
*	236252	236271: gap of unknown length
*	236272	240572: contig of 4301 bp in length
*	240573	240592: gap of unknown length
*	240593	242789: contig of 2197 bp in length
*	242790	242809: gap of unknown length
*	242810	244695: contig of 1886 bp in length
*	244696	244715: gap of unknown length
*	244716	246969: contig of 2254 bp in length
*	246970	246989: gap of unknown length
*	246990	247870: contig of 881 bp in length
*	247871	247890: gap of unknown length
*	247891	248193: contig of 303 bp in length
*	248194	248213: gap of unknown length
*	248214	250122: contig of 1909 bp in length
*	250123	250142: gap of unknown length
*	250143	251367: contig of 1225 bp in length
*	251368	251387: gap of unknown length
*	251388	252749: contig of 1362 bp in length
*	252750	252769: gap of unknown length
*	252770	253640: contig of 871 bp in length
*	253641	253660: gap of unknown length
*	253661	255032: contig of 1372 bp in length
*	255033	255052: gap of unknown length
*	255053	257333: contig of 2281 bp in length
*	257334	257353: gap of unknown length
*	257354	259409: contig of 2056 bp in length
*	259410	259429: gap of unknown length
*	259430	260459: contig of 1030 bp in length
*	260460	260479: gap of unknown length
*	260480	260829: contig of 350 bp in length
*	260830	260849: gap of unknown length
*	260850	261661: contig of 812 bp in length
*	261662	261681: gap of unknown length
*	261682	262532: contig of 851 bp in length
*	262533	262552: gap of unknown length
*	262553	263351: contig of 799 bp in length
*	263352	263371: gap of unknown length
*	263372	264217: contig of 846 bp in length
*	264218	264237: gap of unknown length
*	264238	265205: contig of 968 bp in length
*	265206	265225: gap of unknown length
*	265226	266270: contig of 1045 bp in length
*	266271	266290: gap of unknown length
*	266291	266749: contig of 459 bp in length
*	266750	266769: gap of unknown length
*	266770	267718: contig of 949 bp in length
*	267719	267738: gap of unknown length
*	267739	268576: contig of 838 bp in length
*	268577	268596: gap of unknown length
*	268597	270010: contig of 1414 bp in length
*	270011	270030: gap of unknown length
*	270031	271654: contig of 1624 bp in length
*	271655	271674: gap of unknown length
*	271675	272545: contig of 871 bp in length.

FEATURES Location/Qualifiers

source	1. .272545
	/organism="Mus musculus"
	/strain="C57BL6/J"
	/db_xref="taxon:10090"
	/clone="RP23-33H13"
	/sex="male"
misc_feature	1. .30267
	/note="assembly_name:Contig108"
misc_feature	30288. .58400
	/note="assembly_name:Contig107 clone_end:SP6 vector_side:left"
misc_feature	58421. .88864
	/note="assembly_name:Contig106"
misc_feature	88885. .109889
	/note="assembly_name:Contig105"
misc_feature	109910. .125133
	/note="assembly_name:Contig104 clone_end:T7 vector_side:right"
misc_feature	125154. .139549
	/note="assembly_name:Contig103"
misc_feature	139570. .153410
	/note="assembly_name:Contig102"
misc_feature	153431. .162567
	/note="assembly_name:Contig101"
misc_feature	162588. .171044
	/note="assembly_name:Contig100"
misc_feature	171065. .179304
	/note="assembly_name:Contig99"
misc_feature	179325. .187059
	/note="assembly_name:Contig98"
misc_feature	187080. .197598
	/note="assembly_name:Contig97"
misc_feature	197619. .205966
	/note="assembly_name:Contig96"
misc_feature	205987. .212449
	/note="assembly_name:Contig95"
misc_feature	212470. .217752
	/note="assembly_name:Contig94"
misc_feature	217773. .223943
	/note="assembly_name:Contig93"
misc_feature	223964. .230362
	/note="assembly_name:Contig92"
misc_feature	230383. .236251
	/note="assembly_name:Contig91"

```
     misc_feature    236272. .240572
                     /note="assembly_name:Contig90"
     misc_feature    240593. .242789
                     /note="assembly_name:Contig89"
     misc_feature    242810. .244695
                     /note="assembly_name:Contig88"
     misc_feature    244716. .246969
                     /note="assembly_name:Contig87"
     misc_feature    246990. .247870
                     /note="assembly_name:Contig86"
     misc_feature    247891. .248193
                     /note="assembly_name:Contig85"
     misc_feature    248214. .250122
                     /note="assembly_name:Contig84"

  Query Match             100.0%;  Score 64;  DB 2;  Length 272545;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-11;
  Matches   64;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  257151 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 257210

Qy      61 AGCC 64
           ||||
Db  257211 AGCC 257214


RESULT 10
AP002566/c
LOCUS       AP002566               280900 bp    DNA     linear   BCT 07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 17/20.
ACCESSION   AP002566 BA000007
VERSION     AP002566.1  GI:13363930
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 280900)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .280900
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            complement(143. .616)
                     /gene="ECs4456"
     CDS             complement(143. .616)
                     /gene="ECs4456"
                     /note="similar to YIAI_ECOLI YIAI_ECOLI gi|2367245 percent
                     identity 98 in 157 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB37879.1"
                     /db_xref="GI:13363931"
                     /translation="MNRFIIADATKCIGCRTCEVACAVSHQENQDCAALSPDEFISRI
                     RVIKDHSWTTAVACHQCEDAPCANVCPVDAISREHGHIFVEQTRCIGCKSCMLACPFG
                     AMEVVSSRKKARAIKCDLCWHRETGPACVEACPTKALQCMDVEKVQRHRLRQQPV"
     gene            complement(860. .1675)
                     /gene="ECs4457"
     CDS             complement(860. .1675)
                     /gene="ECs4457"
                     /note="probable transcriptional regulator,weakly similar
                     to hypothetical transcriptional regulator YisR [Bacillus
                     subtilis] gi|3123306|sp|P40331 percent identity 24 in 276
                     aa"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /protein_id="BAB37880.1"
                     /db_xref="GI:13363932"
                     /translation="MLELSMTLPIKVQNGGLFISRGVGRHSARRLESWEIIFVEKGRL
                     TIQEEERIFLVDAGESLLLWPNRQHVGVEDFPADLKFYWLHFEVKAPDNDPRWLTHLS
                     VPQQTKVVDPQALVALFRQFLSEQEKHQRTPALEFIVLLILQQLTRDARLDEQADGAG
                     TTLAWKAQKLIRTQYHRPLSTALLARELYCNADYLGRVYRRVFHLTLTEAIHRQRVRE
                     AEKLLISEARSLTEVAALCGFNDVGYFRQIFRKHTGLTPAAWKRRYSKEHINS"
     gene            1901. .3301
                     /gene="ECs4458"
     CDS             1901. .3301
                     /gene="ECs4458"
                     /note="probable permease,similar to hypothetical proteins
                     [Salmonella typhimurium] gi|7442781|pir||C65167 percent
                     identity 37 in 444 aa,also similar to transport proteins
                     (putative symporters) e.g. YicJ [Escherichia coli (K-12)]
                     gi|2851421|sp|P31435|YICJ_EC"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative permease"
                     /protein_id="BAB37881.1"
                     /db_xref="GI:13363933"
                     /translation="MTSTPITTADIAKGKIDDALSVREKIGYGLGDAGGTVITCLIMN
                     FLTFFYTDVFGLTPALVGTLFIALRVFDAISDPVMGVIADRTQSRWGRFRPWQLWIAI
```

PIGIIGILTFTVPDASMGVKIAWAFGTYLLLSVGYTAINVPYCARINTMTTRHNEVIS
CQSWRFVLCGVAGFLVSVGLPWMVALFGQGNAARGYQLGVGVLCAIAVVMFLCCFFWV
RERVPLSTMGKFTLREHLAGLRNNDQLLLMLVMSFLLINVFNIRGGGYMYFITYVLQG
STGYTSLFFTMVTFASIIGSVIVSPLTRRFDTVKIYYYTNLLLAALAVLMWFLPSGPA
YQTLWLAVILGNGVILGFTLPLHFSLMAFADDYGEWKTHVRSSGMNFAFNLFFIKLAW
ASSAGIISLLFIFVAYQPGVENQTASSLGGIAAMETLLPALFHLLLAGAIRFCKLNNP
MMSRIATDLRQRHVQP"

gene 3312. .5282

CDS 3312. .5282
/gene="ECs4459"
/note="unknown,similar to hypothetical protein [Thermotoga maritima] gi|7452109|pir||F72395 percent identity 37 in 635 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37882.1"
/db_xref="GI:13363934"
/translation="MNISEVDLHKLTVSDPFLGQYQQLVRDVVIPYQWDALNDRIPEA
EPSHAIENFRIAAGLQEGEFYGMVFQDSDVAKWLEAVAWSLCQKPDAELEKTADEVIE
LIASAQCEDGYLNTYFTVKAPEERWSNLAECHELYCAGHLIEAGVAFFQATGKRRLLG
VVCRLADHIDSVFGPDESKLHGYPGHPEIELALMRLYEVTEEPRYLALTNYFVEQRGA
QPHYYDQEYEKRGQTSHWHTYGPAWMVKDKAYSQAHLPLAQQQTAIGHAVRFVYLMTG
VAHLARLSHDDSKRQDCLRLWNNMAQRQLYITGGIGSQSSGEAFSSDYDLPNDTVYAE
SCASIGLMMFARRMLEMEGDSQYADVMERALYNTVLGGMALDGKHFFYVNPLEVHPKS
LKFNHIYDHVKPIRQRWFGCACCPPNIARVLTSIGHYLYTPREDALYINIYAGNSMEV
PVENGTLRLRVSGNYPWQEQVTIAVESPQPVRHTLALRLPDWCTQPQIILNGEEVEQD
IRKGYLHITREWQEGDTLNLTLPMPVRRVYGNPLVRHVAGKVAIQRGPLVYCLEKADN
GESLHNLWLPTDAPFTTFEGKGLFSHKILIQAPGYRYEQSNPEQQPLWHYDSAPAKRQ
TQTLTFIPWFSWANRGEGEMRIWVNEEKHCHP"

gene complement(5412. .6152)

CDS complement(5412. .6152)
/gene="ECs4460"
/note="similar to YIAT_ECOLI gi|1790010 percent identity 99 in 246 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative outer membrane protein"
/protein_id="BAB37883.1"
/db_xref="GI:13363935"
/translation="MLINRNIVALFALPFMASATASELSIGAGAAYNESPYRGYNENT
KAIPLISYEGDSFYVRQTTLGFILSQSEKNELSLTASWMPLEFDPADNDDYAMQQLDK
RDSTAMAGVAWYHHERWGTVKASAAADVLDNSNGWVGELSVFHKMQIGRLSLTPALGV
LYYDENFSDYYYGISESESRRSGLASYSAQDAWVPYVSLTAKYPIGEHVVLMASAGYS
ELPEEITDSPMIDRNESFTFVTGVSWRF"

gene 6276. .7250

CDS 6276. .7250
/gene="ECs4461"
/note="similar to YIAU_ECOLI gi|1790011 percent identity 99 in 324 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transcriptional regulator LYSR-type"
/protein_id="BAB37884.1"
/db_xref="GI:13363936"
/translation="MTKLQLKYRELKIISVIAASENISHAATVLGIAQANVSKYLADF
ESKVGLKVFDRTTRQLMLTLFGTALLPYINDMLDRNEQLNNFIADYKHEKRGRVTIYA
PTGIITYLSKHVIDKIKDIGDITLSLKTCNLERNAFYEGVEFPDDCDVLISYAPPKDE
SLVASFITQYAVTAYASQRYLEKHPISRPDELEHHSCILIDSMMIDDANIWRFNVAGS
KEVRDYRVKGNYVCDNTQSALELARNHLGIVFAPDKSVQSDLQDGTLVPCFQHPYEWW
LDLVAIFRKREYQPWRVQYVLDEMLREIRHQLAQSQQLRPEQAAESED"

gene complement(7247. .8383)

CDS complement(7247. .8383)
/gene="ECs4462"
/note="similar to YIAV_ECOLI gi|1790012 percent identity 99 in 378 aa, similar to YIBH_ECOLI gi|1790024 percent identity 76 in 378 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative membrane protein"
/protein_id="BAB37885.1"
/db_xref="GI:13363937"
/translation="MDLLIILTYVAFAWAMFKIFKIPVNKWTIPTAALGGIFIVSGLI
LLMNYNHPYTFKAQKAVISIPVVPQVTGVVIEVTDKKNTLIKKGEVLFRLDPTRYQAR
VDRLMADIVTAEHKQRALGAELDEMAANTQKAKATRDKFAKEYQRYARGSQAKVNPFS
ERDIDVARQNYLAQEASVKSSAAEQKQIQSQLDSLVLGEHSQIASLKAQLAEAKYNLE

Query Match 100.0%; Score 64; DB 1; Length 280900;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 171511 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 171452

Qy 61 AGCC 64
||||
Db 171451 AGCC 171448

RESULT 11
AL627280
LOCUS AL627280 245050 bp DNA linear BCT 06-JUN-2002
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome; segment 16/20.
ACCESSION AL627280 AL513382
VERSION AL627280.1 GI:16504729
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi.
ORGANISM Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 245050)
AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T., Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N., Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P., Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.
TITLE Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18
JOURNAL Nature 413 (6858), 848-852 (2001)
MEDLINE 21534947
PUBMED 11677608
REFERENCE 2 (bases 1 to 245050)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES Location/Qualifiers
source 1. .245050
/organism="Salmonella enterica subsp. enterica serovar Typhi"
/strain="CT18"
/db_xref="taxon:90370"
gene 319. .1728
/gene="STY3874"
/note="glnA"
CDS 319. .1728
/gene="STY3874"

```
                     /note="Orthologue of E. coli glnA (GLNA_ECOLI); Fasta hit
                     to GLNA_ECOLI (468 aa), 98% identity in 468 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamine synthetase"
                     /protein_id="CAD03093.1"
                     /db_xref="GI:16504730"
                     /db_xref="SWISS-PROT:P06201"
                     /translation="MSAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFF
                     EEGKMFDGSSIGGWKGINESDMVLMPDASTAVIDPFFADSTLIIRCDILEPGTLQGYD
                     RDPRSIAKRAEDYLRATGIADTVLFGPEPEFFLFDDIRFGASISGSHVAIDDIEGAWN
                     SSTKYEGGNKGHRPGVKGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAHHHEVATAG
                     QNEVATRFNTMTKKADEIQIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLA
                     KNGTNLFSGDKYAGLSEQALYYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVMLA
                     YSARNRSASIRIPVVASPKARRIEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPGEA
                     MDKNLYDLPPEEAKEIPQVAGSLEEALNALDLDREFLKAGGVFTDEAIDAYIALRREE
                     DDRVRMTPHPVEFELYYSV"
misc_feature         364. .1434
                     /gene="STY3874"
                     /note="Pfam match to entry PF00120 gln-synt, Glutamine
                     synthetase, score 838.00, E-value 2.4e-266"
misc_feature         466. .522
                     /gene="STY3874"
                     /note="PS00180 Glutamine synthetase signature 1"
misc_feature         1093. .1140
                     /gene="STY3874"
                     /note="PS00181 Glutamine synthetase putative ATP-binding
                     region signature"
misc_feature         1474. .1512
                     /gene="STY3874"
                     /note="PS00182 Glutamine synthetase class-I adenylation
                     site"
gene                 2003. .3052
                     /gene="STY3875"
                     /note="glnL"
CDS                  2003. .3052
                     /gene="STY3875"
                     /note="Orthologue of E. coli glnL (NTRB_ECOLI); Fasta hit
                     to NTRB_ECOLI (349 aa), 93% identity in 349 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="Two-component system sensory histidine kinase"
                     /protein_id="CAD03094.1"
                     /db_xref="GI:16504731"
                     /db_xref="SWISS-PROT:P41788"
                     /translation="MASGIQPDAGQILNSLINSVLVVDDALAIHYANPAAQQLLAQSS
                     RKLFGTPLPELLSYFSLNIDLMRESLAAGQGFTDNEVTLVIDSRSHILSLTAQRLPDD
                     FILLEMAPMDNQRRLSQEQLQHAQQIAARDLVRGLAHEIKNPLGGLRGAAQLLSKALP
                     DPALTEYTKVIIEQADRLRNLVDRLLGPQHPGMHITESIHKVAERVVALVSMELPDNV
                     RLIRDYDPSLPELPHDPEQIEQVLLNIVRNALQALGPEGGEITLRTRTAFQLTLHGER
                     YRLAARIDVEDNGPGIPPHLQDTLFYPMVSGREGGTGLGLSIARNLIDQHAGKIEFTS
                     WPGHTEFSVYLPIRK"
misc_feature         2417. .3040
                     /gene="STY3875"
                     /note="Pfam match to entry PF00512 signal, Histidine
                     kinase, score 244.40, E-value 1.6e-69"
gene                 3061. .4470
                     /gene="STY3876"
                     /note="glnG"
CDS                  3061. .4470
                     /gene="STY3876"
                     /note="Fasta hit to HYDG_ECOLI (441 aa), 44% identity in
                     462 aa overlap
                     Fasta hit to YGAA_ECOLI (529 aa), 36% identity in 451 aa
                     overlap
                     Fasta hit to ATOC_ECOLI (461 aa), 43% identity in 468 aa
                     overlap
                     Fasta hit to YFHA_ECOLI (444 aa), 44% identity in 381 aa
                     overlap
                     Orthologue of E. coli glnG (NTRC_ECOLI); Fasta hit to
                     NTRC_ECOLI (469 aa), 96% identity in 469 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="Two-component system, response regulator"
                     /protein_id="CAD03095.1"
                     /db_xref="GI:16504732"
                     /db_xref="SPTREMBL:Q822S3"
                     /translation="MQRGIVWVVDDDSSIRWVLERALAGAGLTCITFENGNEVLAALA
                     SKTPDVLLSDIRMPGMDGLALLKQIKQRHPMLPVIIMTAHSDLDAAVSAYQQGAFDYL
                     PKPFDIDEAVALVERAISHYQEQQQPRNIEVNGPTTDMIGEAPAMQDVFRIIGRLSRS
                     SISVLINGESGTGKELVAHALHRHSPRAKAPFIALNMAAIPKDLIESELFGHEKGAFT
                     GANTIRQGRFEQADGGTLFLDEIGDMPLDVQTRLLRVLADGQFYRVGGYAPVKVDVRI
                     IAATHQNLERRVQEGKFREDLFHRLNVIRIHLPPLRERREDIPRLARHFLQVAARELG
                     VEAKLLHPETETALTRLAWPGNVRQLENTCRWLTVMAAGQEVLTQDLPGELFEASAPD
                     SPSHLPPDSWATLLAQWADRALRSGHQNLLSEAQPELERTLLTTALRHTQGHKQEAAK
                     LLGWGRNTLTRKLKELGME"
misc_feature         3070. .3405
                     /gene="STY3876"
                     /note="Pfam match to entry PF00072 response_reg, Response
                     regulator receiver domain, score 141.50, E-value 1.5e-38"
misc_feature         3478. .4173
                     /gene="STY3876"
                     /note="Pfam match to entry PF00158 sigma54, Sigma-54
                     transcription factors, score 515.80, E-value 3.1e-151"
misc_feature         3550. .3591
                     /gene="STY3876"
                     /note="PS00675 Sigma-54 interaction domain ATP-binding
                     region A signature"
misc_feature         3736. .3783
                     /gene="STY3876"
                     /note="PS00676 Sigma-54 interaction domain ATP-binding
                     region B signature"
misc_feature         4117. .4146
                     /gene="STY3876"
                     /note="PS00688 Sigma-54 interaction domain C-terminal part
                     signature"
gene                 complement(4783. .6156)
                     /gene="STY3877"
                     /note="hemN"
CDS                  complement(4783. .6156)
                     /gene="STY3877"
                     /note="Orthologue of E. coli hemN (HEMN_ECOLI); Fasta hit
                     to HEMN_ECOLI (457 aa), 93% identity in 457 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="oxygen-independent coproporphyrinogen III
                     oxidase"
                     /protein_id="CAD03096.1"
                     /db_xref="GI:16504733"
                     /db_xref="SWISS-PROT:P37129"
                     /translation="MSEQQIDWDLALIQKYNYSGPRYTSYPTALEFSEDFEDAAFLQA
                     VARYPERPLSLYVHIPFCHKLCYFCGCNKIVTRQQHKADQYLDALEQEIRHRAPLFAD
                     RHVSQLHWGGGTPTYLNKAQISRLMTLLRENFHFNTDAEISIEVDPREIELDVLDHLR
                     AEGFNRLSMGVQDFNKEVQRLVNREQDEEFIFALLNHARDIGFTSTNIDLIYGLPKQT
                     PESFAFTLKRVTELNPDRLSVFNYAHLPTLFAAQRKIKDADLPSAQQKLDILQETIVS
                     LTQAGYQFIGMDHFARPDDELAVAQREGVLHRNFQGYTTQGDTDLLGMGVSAISMIGD
                     GYMQNQKELKRYYQQVDERGNALWRGITLTRDDCIRRDVIKALICNFRLDFNAVEQQW
                     GLHFAEYFAEDLQLLSPLAKDGLVDISEKGIQVTAKGRLLIRNICMCFDAYLRQKARM
                     QQFSRVI"
gene                 complement(6345. .6860)
                     /gene="STY3878"
                     /note="yihI"
CDS                  complement(6345. .6860)
                     /gene="STY3878"
                     /note="Orthologue of E. coli yihI (YIHI_ECOLI); Fasta hit
                     to YIHI_ECOLI (169 aa), 85% identity in 169 aa overlap.
                     Note this protein is extremely hydrophyllic"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD03097.1"
                     /db_xref="GI:16504734"
                     /db_xref="SWISS-PROT:P37130"
                     /translation="MKKPTSAPRSKAFGKQRRKTREELNQEARDRKRLKKHRGHAPGS
                     RAAGGNSASGGGNQNQQKDPRIGSKTPVPLGVTEKVTQQHKPKSEKPMLSPQAELDLL
                     ETDERLDALLERLEAGETLSAEDQAWVDAKLDRIDELMQKLGLSYDDDEEDDEEDEKQ
```

```
                     EDMMRLLRGGN"
     gene            complement(6979. .7203)
                     /gene="STY3879"
     CDS             complement(6979. .7203)
                     /gene="STY3879"
                     /note="no significant database hits. Note this protein is
                     extremely hydrophyllic"
                     /codon_start=1

  Query Match              61.9%;  Score 39.6;  DB 1;  Length 245050;
  Best Local Similarity    77.4%;  Pred. No. 0.0031;
  Matches   48;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           |||||||||||| | | | ||  |||||||||| | | | |||||||||||| ||   |||
Db  110239 CAAAAAAAAGTCGGATTAAAATACTTGAAAAGTACGTATTCAGACCCATTTATAAGACGT 110298

Qy      61 AG 62
           ||
Db  110299 AG 110300


RESULT 12
AE008877/c
LOCUS       AE008877               22492 bp    DNA     linear   BCT 31-JUL-2002
DEFINITION  Salmonella typhimurium LT2, section 181 of 220 of the complete
            genome.
ACCESSION   AE008877 AE006468
VERSION     AE008877.1  GI:16422367
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2.
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 22492)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22492)
  AUTHORS   The Salmonella typhimurium Genome Sequencing Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .22492
                     /organism="Salmonella typhimurium LT2"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            complement(78. .278)
                     /gene="ivbL"
                     /note="STM3797"
     CDS             complement(78. .176)
                     /gene="ivbL"
                     /note="similar to E. coli ilvB operon leader peptide
                     (AAC76695.1); Blastp hit to AAC76695.1 (32 aa), 84%
                     identity in aa 1 - 32"
                     /codon_start=1
                     /transl_table=11
                     /product="ilvB operon leader peptide"
                     /protein_id="AAL22655.1"
                     /db_xref="GI:16422368"
                     /translation="MNPSMLNATLLTTAPSRAVVVVRVVVVVGNAP"
     RBS             complement(181. .186)
                     /gene="ivbL"
                     /note="putative RBS for ivbL; RegulonDB:STMS1H003766"
     -10_signal      complement(254. .262)
                     /gene="ivbL"
                     /note="putative -10_signal for ivbL;
                     RegulonDB:STMLTH004613"
     protein_bind    complement(264. .292)
                     /note="putative binding site for CRP, RegulonDB:
                     STMS1H000038"
                     /bound_moiety="CRP"
     -35_signal      complement(270. .278)
                     /gene="ivbL"
                     /note="putative -35_signal for ivbL;
                     RegulonDB:STMLTH004613"
     gene            699. .813
                     /gene="ysdA"
                     /note="STM3797A"
     CDS             join(699. .790,789. .813)
                     /gene="ysdA"
                     /note="ribosomal slippage"
                     /codon_start=1
                     /transl_table=11
                     /product="YsdA"
                     /protein_id="AAM89474.1"
                     /db_xref="GI:22025623"
                     /translation="MSTRTLFLAKGIACATAGNRQRKGTSAGSAPLNHRRRV"
     gene            810. .899
                     /gene="ysdB"
                     /note="STM3796B"
     CDS             810. .899
                     /gene="ysdB"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved protein in the LexA regulon"
                     /protein_id="AAL22656.1"
                     /db_xref="GI:16422369"
                     /translation="MSVVDITILILKLIVAALQLLDAVLKYLK"
     gene            992. .1852
                     /gene="STM3796A"
     CDS             992. .1852
                     /gene="STM3796A"
                     /codon_start=1
                     /transl_table=11
                     /product="integral membrane protein"
                     /protein_id="AAL22657.1"
                     /db_xref="GI:16422370"
                     /translation="MTPSTTPDAMTLSVFCILLFAALLHASWNAIVKAGNDKLYAAIG
                     VSGSAAVMALILLPFSPQPAHASIPFLAASTALQVVYTVLVAKTYQVSDMSQTYPLMR
                     GTAPLLVALISVLFLGDSLSSLAWVGIAVICMAILGMACNGRASSQRGVVLALTNACF
                     IAGYTLVDGTGVRLSETALGYTLWSFFLNGACLLTWAMIARRREASRYLAQQWKKGIF
```

```
                GGIGTMGSYGLALWAMTQAPLAVVAALRETSILFGALIAWLLLKEKVAGLRLVAAGGI
                ALGAILLRLS"
     gene       2055. .3239
                /gene="emrD"
                /note="STM3798"
     CDS        2055. .3239
                /gene="emrD"
                /note="similar to E. coli 2-module integral membrane pump;
                multidrug resistance (AAC76696.1); Blastp hit to
                AAC76696.1 (396.aa), 92% identity in aa 3 - 396"
                /codon_start=1
                /transl_table=11
                /product="MFS family multidrug tranport protein"
                /protein_id="AAL22658.1"
                /db_xref="GI:16422371"
                /translation="MKRQRNVNLLLMLVLLVAVGQMAQTIYIPAIADMAQALNVREGA
                VQSVMAAYLLTYGVSQLFYGPLSDRVGRRPVILVGMSIFMVATLFAMTTHSLTVLIAA
                SAMQGMGTGVGGVMARTLPRDLYEGTQLRHANSLLNMGILVSPLLAPLIGSLLDTLWN
                WRACYAFLLVLCAGVTFSMARWMPETRPAGAPRTRLIASYKTLFGNGAFNCYLLMLIG
                GLAGIAVFEACSGVLMGAVLGLSSMVVSILFILPIPAAFFGAWFAGRPNKRFSTLMWQ
                SVICCLLAGLMMWIPGWFGVMNVWTLLIPAALFFFGAGMLFPLATSGAMEPFPFLAGT
                AGALVGGLQNIGSGVLAWLSAMLPQTGQGSLGLLMTLMGLLILACWLPLASRISHQGQ
                TV"
     gene       complement(3166. .4251)
                /gene="STM3799"
     CDS        complement(3166. .4251)
                /gene="STM3799"
                /codon_start=1
                /transl_table=11
                /product="putative periplasmic protein"
                /protein_id="AAL22659.1"
                /db_xref="GI:16422372"
                /translation="MVKILGGVVFKPLIASLMLTSAVVYAKPMPLTAARYAQQLGVGM
                DVDWARTERGIREFDPLVVRDFKAKGLTHVRIRVAGAPTEARLIHLRKLVEACEYYGV
                IPIIAYQADAYKTDPSASHEKELINWWSVVARYFGQTSPLLGFDLIYEPADKLNHNMA
                SLNRVYDKTIRLIHAIDPQRMIFVAPRMRAAPEDLSALKLPAQSQNYVLAEWHIFPWG
                PLKSGGKYPWTSGTAAEKAAIRARINAAVRWQHKTGHASWVGGWAPGESIKMTPVASQ
                FAFARFMACELKKAHIPYAINADTQFYDGEEGAWRPAQESLLNAMIAPECETPGKTPG
                EGNVKPSAPDAISVTPAAASTPGSAIP"
     gene       complement(4315. .5247)
                /gene="dsdC"
                /note="STM3800"
     CDS        complement(4315. .5238)
                /gene="dsdC"
                /note="similar to E. coli D-serine dehydratase (deaminase)
                transcriptional activator (AAC75423.1); Blastp hit to
                AAC75423.1 (311 aa), 90% identity in aa 1 - 307"
                /codon_start=1
                /transl_table=11
                /product="LysR family transcriptional activator"
                /protein_id="AAL22660.1"
                /db_xref="GI:16422373"
                /translation="MEPLRDTRNRLLNGWQLSKLYTFEVAARHQSFALAADELSLSPS
                AVSHRINQLEAELGIQLFVRSHRKVELTHEGKRVFWALKSSLDTLNQEILDIKNQELS
                GTLTVYSRPSIAQCWLVPALGDFTRRYPSISLTMLTGNDNVNLQRAGIDLAIYFDDAP
                SAQLTHHFLMDEAILPVCSPDYARRFALQDNLNNLRHCTLLHDRQAWSNDSGTDEWHS
                WAQHFAVNLPDSSGIGFDRSDLAVIAAMNHIGVAMGRKRLVQKRLDSGELIAPFGDMR
                LKCHQHYYVTTLPGRQWPKIEAFIRWLQEQV"
     RBS        complement(5242. .5247)
                /gene="dsdC"
                /note="putative RBS for dsdC; RegulonDB:STMS1H003767"
     protein_bind 5316. .5327
                /note="putative binding site for CRP, RegulonDB:
                STMS1H000044"
                /bound_moiety="CRP"
     protein_bind 5342. .5362
                /note="putative binding site for CRP, RegulonDB:
                STMS1H000090"
                /bound_moiety="CRP"
     gene       5361. .6801
                /gene="dsdX"
                /note="STM3801"
     -35_signal 5361. .5369
                /gene="dsdX"
                /note="putative -35_signal for dsdX;
                RegulonDB:STMLTH004614"
     -10_signal 5388. .5396
                /gene="dsdX"

  Query Match           59.4%;  Score 38;  DB 1;  Length 22492;
  Best Local Similarity  75.8%;  Pred. No. 0.016;
  Matches   47;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           |||||||||| | | | | ||  |||||||||| | | | ||||||||||| ||   |||
Db   12852 CAAAAAAAAGCCGGATTAAAATACTTGAAAAGTACGTATTCAGACCCATTTATAAGACGT 12793

Qy      61 AG 62
           ||
Db   12792 AG 12791


RESULT 13
AC002459/c
LOCUS       AC002459               25869 bp    DNA     linear   PRI 21-FEB-2002
DEFINITION  Homo sapiens BAC clone CTB-188O5 from 7q22-q31, complete sequence.
ACCESSION   AC002459
VERSION     AC002459.2  GI:18855173
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 25869)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 25869)
  AUTHORS   Madsen,C., Graves,T. and Cofman,M.
  TITLE     The sequence of Homo sapiens BAC clone CTB-188O5
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 25869)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 25869)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 25869)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 25869)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 21, 2002 this sequence version replaced gi:2337870.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_RG188O05
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
```

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
Clone CTB-188O5 is from the first release of the human BAC library CITB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is CTA-313E3, 200 bp overlap. Actual start of this clone is at base position 122698 of CTA-313E3; actual end is at base position 25869 of CTB-188O5.

FEATURES Location/Qualifiers

source	1. .25869
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q22-q31"
	/clone="CTB-188O5"
	/clone_lib="CITB-978SK-B"
repeat_region	2. .466
	/rpt_family="L1"
repeat_region	459. .520
	/rpt_family="L1"
repeat_region	521. .826
	/rpt_family="Alu"
repeat_region	803. .826
	/rpt_family="(A)n"
repeat_region	827. .1564
	/rpt_family="L1"
repeat_region	1622. .1960
	/rpt_family="AcHobo"
repeat_region	1961. .2274
	/rpt_family="ERVL"
repeat_region	2431. .2548
	/rpt_family="AcHobo"
repeat_region	2442. .2469
	/rpt_family="AT_rich"
repeat_region	2549. .2870
	/rpt_family="L1"
repeat_region	2871. .2936
	/rpt_family="AcHobo"
repeat_region	3128. .3424
	/rpt_family="Alu"
repeat_region	4043. .4236
	/rpt_family="MIR"
repeat_region	4209. .4239
	/rpt_family="AT_rich"
repeat_region	6203. .6523
	/rpt_family="Alu"
repeat_region	6322. .6350
	/rpt_family="(A)n"
misc_feature	6709. .6824
	/note="match to EST BF592945 (NID:g11685269)"
repeat_region	7093. .7156
	/rpt_family="L1"
repeat_region	8790. .9147
	/rpt_family="MaLR"
repeat_region	10502. .10744
	/rpt_family="MaLR"
repeat_region	11131. .11158
	/rpt_family="AT_rich"
repeat_region	11322. .11481
	/rpt_family="MIR"
repeat_region	13163. .13191
	/rpt_family="(TTTA)n"
repeat_region	13279. .13543
	/rpt_family="L2"
repeat_region	13629. .13771
	/rpt_family="L1"
repeat_region	13829. .14233
	/rpt_family="L1"
repeat_region	14234. .14767
	/rpt_family="L1"
repeat_region	14691. .14714
	/rpt_family="AT_rich"
repeat_region	14772. .16354
	/rpt_family="L1"
repeat_region	16355. .16869
	/rpt_family="L1"
repeat_region	16889. .16974
	/rpt_family="L1"
repeat_region	16983. .17258
	/rpt_family="Alu"
repeat_region	17382. .17407
	/rpt_family="AT_rich"
repeat_region	17582. .17731
	/rpt_family="L2"
repeat_region	18560. .18705
	/rpt_family="L1"
misc_feature	20341. .20633
	/note="match to EST BG773033 (NID:g14083686)"
repeat_region	20539. .20745
	/rpt_family="MIR"
repeat_region	21127. .21603
	/rpt_family="ERV1"
repeat_region	21608. .21914
	/rpt_family="Alu"
repeat_region	21917. .22190
	/rpt_family="L1"
repeat_region	22199. .22218
	/rpt_family="(TTTTG)n"
repeat_region	22200. .22371
	/rpt_family="Alu"
repeat_region	22421. .22977
	/rpt_family="ERVL"
repeat_region	23267. .23429
	/rpt_family="Alu"
repeat_region	23410. .23430
	/rpt_family="AT_rich"
repeat_region	23462. .23519
	/rpt_family="MaLR"
repeat_region	23600. .23748
	/rpt_family="MaLR"
repeat_region	23749. .24334
	/rpt_family="ERV1"
repeat_region	24335. .24548
	/rpt_family="MaLR"
repeat_region	24600. .24700
	/rpt_family="MaLR"
repeat_region	24975. .25812
	/rpt_family="L2"

BASE COUNT 7711 a 5089 c 5453 g 7616 t
ORIGIN

```
  Query Match             45.0%;  Score 28.8;  DB 9;  Length 25869;
  Best Local Similarity   65.6%;  Pred. No. 18;
  Matches   42;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           |||||| |  | | ||   ||| ||| || |   |||  || | ||||| ||| ||||
Db   19526 CAAAAATACTTGCCCTTTGAAGTCTTAAATATACCATCCCCCGCCCCATGTTTCCATCTA 19467

Qy      61 AGCC 64
           ||||
Db   19466 AGCC 19463


RESULT 14
AC129514/c
LOCUS       AC129514               66873 bp    DNA     linear   HTG 30-JUL-2002
DEFINITION  Homo sapiens chromosome 18 clone RP11-61K3 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC129514
VERSION     AC129514.1  GI:22004327
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 66873)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-61K3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 66873)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L27809
                Center clone name: 61_K_3
            --------------.
            * NOTE: This record contains 81 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *      1      729: contig of 729 bp in length
            *    730  829: gap of      100 bp
            *    830     1552: contig of 723 bp in length
            *   1553  1652: gap of      100 bp
            *   1653     2377: contig of 725 bp in length
            *   2378  2477: gap of      100 bp
            *   2478     3209: contig of 732 bp in length
            *   3210  3309: gap of      100 bp
            *   3310     4045: contig of 736 bp in length
            *   4046  4145: gap of      100 bp
            *   4146     4868: contig of 723 bp in length
            *   4869  4968: gap of      100 bp
            *   4969     5703: contig of 735 bp in length
            *   5704  5803: gap of      100 bp
            *   5804     6549: contig of 746 bp in length
            *   6550  6649: gap of      100 bp
            *   6650     7387: contig of 738 bp in length
            *   7388  7487: gap of      100 bp
            *   7488     8212: contig of 725 bp in length
            *   8213  8312: gap of      100 bp
            *   8313     9033: contig of 721 bp in length
            *   9034  9133: gap of      100 bp
            *   9134     9842: contig of 709 bp in length
            *   9843  9942: gap of      100 bp
            *   9943    10672: contig of 730 bp in length
            *  10673 10772: gap of      100 bp
            *  10773    11514: contig of 742 bp in length
            *  11515 11614: gap of      100 bp
            *  11615    12328: contig of 714 bp in length
            *  12329 12428: gap of      100 bp
            *  12429    13168: contig of 740 bp in length
            *  13169 13268: gap of      100 bp
            *  13269    13992: contig of 724 bp in length
            *  13993 14092: gap of      100 bp
            *  14093    14819: contig of 727 bp in length
            *  14820 14919: gap of      100 bp
            *  14920    15646: contig of 727 bp in length
            *  15647 15746: gap of      100 bp
            *  15747    16470: contig of 724 bp in length
            *  16471 16570: gap of      100 bp
            *  16571    17299: contig of 729 bp in length
            *  17300 17399: gap of      100 bp
            *  17400    18129: contig of 730 bp in length
            *  18130 18229: gap of      100 bp
            *  18230    18958: contig of 729 bp in length
            *  18959 19058: gap of      100 bp
            *  19059    19774: contig of 716 bp in length
            *  19775 19874: gap of      100 bp
            *  19875    20611: contig of 737 bp in length
            *  20612 20711: gap of      100 bp
            *  20712    21448: contig of 737 bp in length
            *  21449 21548: gap of      100 bp
            *  21549    22300: contig of 752 bp in length
            *  22301 22400: gap of      100 bp
            *  22401    23141: contig of 741 bp in length
            *  23142 23241: gap of      100 bp
            *  23242    23966: contig of 725 bp in length
            *  23967 24066: gap of      100 bp
            *  24067    24788: contig of 722 bp in length
            *  24789 24888: gap of      100 bp
            *  24889    25604: contig of 716 bp in length
            *  25605 25704: gap of      100 bp
            *  25705    26430: contig of 726 bp in length
            *  26431 26530: gap of      100 bp
            *  26531    27252: contig of 722 bp in length
            *  27253 27352: gap of      100 bp
            *  27353    28083: contig of 731 bp in length
            *  28084 28183: gap of      100 bp
            *  28184    28887: contig of 704 bp in length
            *  28888 28987: gap of      100 bp
            *  28988    29731: contig of 744 bp in length
```

```
*    29732 29831: gap of      100 bp
*    29832    30559: contig of 728 bp in length
*    30560 30659: gap of      100 bp
*    30660    31392: contig of 733 bp in length
*    31393 31492: gap of      100 bp
*    31493    32226: contig of 734 bp in length
*    32227 32326: gap of      100 bp
*    32327    33050: contig of 724 bp in length
*    33051 33150: gap of      100 bp
*    33151    33879: contig of 729 bp in length
*    33880 33979: gap of      100 bp
*    33980    34712: contig of 733 bp in length
*    34713 34812: gap of      100 bp
*    34813    35539: contig of 727 bp in length
*    35540 35639: gap of      100 bp
*    35640    36246: contig of 607 bp in length
*    36247 36346: gap of      100 bp
*    36347    37074: contig of 728 bp in length
*    37075 37174: gap of      100 bp
*    37175    37904: contig of 730 bp in length
*    37905 38004: gap of      100 bp
*    38005    38741: contig of 737 bp in length
*    38742 38841: gap of      100 bp
*    38842    39581: contig of 740 bp in length
*    39582 39681: gap of      100 bp
*    39682    40421: contig of 740 bp in length
*    40422 40521: gap of      100 bp
*    40522    41255: contig of 734 bp in length
*    41256 41355: gap of      100 bp
*    41356    42098: contig of 743 bp in length
*    42099 42198: gap of      100 bp
*    42199    42922: contig of 724 bp in length
*    42923 43022: gap of      100 bp
*    43023    43743: contig of 721 bp in length
*    43744 43843: gap of      100 bp
*    43844    44568: contig of 725 bp in length
*    44569 44668: gap of      100 bp
*    44669    45393: contig of 725 bp in length
*    45394 45493: gap of      100 bp
*    45494    46212: contig of 719 bp in length
*    46213 46312: gap of      100 bp
*    46313    47052: contig of 740 bp in length
*    47053 47152: gap of      100 bp
*    47153    47875: contig of 723 bp in length
*    47876 47975: gap of      100 bp
*    47976    48676: contig of 701 bp in length
*    48677 48776: gap of      100 bp
*    48777    49503: contig of 727 bp in length
*    49504 49603: gap of      100 bp
*    49604    50318: contig of 715 bp in length
*    50319 50418: gap of      100 bp
*    50419    51149: contig of 731 bp in length
*    51150 51249: gap of      100 bp
*    51250    51972: contig of 723 bp in length
*    51973 52072: gap of      100 bp
*    52073    52808: contig of 736 bp in length
*    52809 52908: gap of      100 bp
*    52909    53649: contig of 741 bp in length
*    53650 53749: gap of      100 bp
*    53750    54485: contig of 736 bp in length
*    54486 54585: gap of      100 bp
*    54586    55317: contig of 732 bp in length
*    55318 55417: gap of      100 bp
*    55418    56154: contig of 737 bp in length
*    56155 56254: gap of      100 bp
*    56255    56994: contig of 740 bp in length
*    56995 57094: gap of      100 bp
*    57095    57824: contig of 730 bp in length
*    57825 57924: gap of      100 bp
*    57925    58637: contig of 713 bp in length

Query Match              42.8%;  Score 27.4;  DB 2;  Length 66873;
Best Local Similarity    69.8%;  Pred. No. 45;

  Matches   37;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       5 AAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACAT 57
           |||  ||   |   ||  ||  ||   ||||||||||||     |||||||||  |  ||  ||
Db   13562 AAACAGATCAATGGTATGGAATGAAAAGTTCAGAAACAGACCCATATATATAT 13510

RESULT 15
AC099848
LOCUS       AC099848               132288 bp    DNA     linear   PRI 14-JUN-2002
DEFINITION  Homo sapiens chromosome 18, clone RP11-856F24, complete sequence.
ACCESSION   AC099848
VERSION     AC099848.2  GI:21426215
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 132288)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-856F24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 132288)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 132288)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 14, 2002 this sequence version replaced gi:17047211.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L13216
                Center clone name: 856_F_24
            --------------

            Only the middle 132.60 kb of this clone are being submitted.
            The remainder of the clone is overlapped by either accession number
            AC025679 [WICGR L8287] or AC055840 [WICGR L9688].
FEATURES              Location/Qualifiers
     source           1. .132288
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="18"
                      /map="18"
                      /clone="RP11-856F24"
                      /clone_lib="RPCI-11 Human Male BAC"
     repeat_region    complement(654. .855)
                      /rpt_family="MIR"
     repeat_region    complement(908. .1042)
                      /rpt_family="MIR"
     repeat_region    1547. .1626
                      /rpt_family="L2"
     repeat_region    1967. .3121
                      /rpt_family="L2"
     repeat_region    complement(3122. .3419)
                      /rpt_family="AluSx"
     repeat_region    3420. .3430
                      /rpt_family="L2"
     repeat_region    complement(3431. .3720)
                      /rpt_family="AluJo"
     repeat_region    3721. .3837
                      /rpt_family="L2"
     repeat_region    5623. .5762
                      /rpt_family="L3"
     repeat_region    complement(7589. .7846)
                      /rpt_family="AluSp"
     repeat_region    7998. .8241
                      /rpt_family="MER41B"
     repeat_region    8577. .8867
                      /rpt_family="AluSg"
     repeat_region    10208. .10340
                      /rpt_family="L1ME4A"
     repeat_region    10393. .10504
                      /rpt_family="L1ME4A"
     repeat_region    complement(10575. .10688)
                      /rpt_family="FLAM_A"
     repeat_region    10997. .11228
                      /rpt_family="L1MCc"
     repeat_region    11229. .11562
                      /rpt_family="AluJb"
     repeat_region    11563. .11743
                      /rpt_family="L1MCc"
     repeat_region    complement(11995. .12028)
                      /rpt_family="MIR"
     repeat_region    12116. .12598
                      /rpt_family="L3"
     repeat_region    15397. .15499
                      /rpt_family="CT-rich"
     repeat_region    complement(15501. .15869)
                      /rpt_family="MSTD"
     repeat_region    16858. .17217
                      /rpt_family="THE1B"
     repeat_region    17664. .17755
                      /rpt_family="MIR"
     repeat_region    17756. .18059
                      /rpt_family="AluSx"
     repeat_region    18060. .18105
                      /rpt_family="MIR"
     repeat_region    18145. .18201
                      /rpt_family="(CATATA)n"
     repeat_region    complement(18250. .18535)
                      /rpt_family="AluSg"
     repeat_region    complement(19762. .19936)
                      /rpt_family="L1ME4A"
     repeat_region    20086. .20400
                      /rpt_family="AluSx"
     repeat_region    20979. .21177
                      /rpt_family="L1MC/D"
     repeat_region    complement(21197. .21372)
                      /rpt_family="MIR"
     repeat_region    complement(21742. .22031)
                      /rpt_family="AluJb"
     repeat_region    22119. .22186
                      /rpt_family="L1ME4A"
     repeat_region    23162. .23305
                      /rpt_family="MIR"
     repeat_region    complement(23415. .26247)
                      /rpt_family="L1M4"
     repeat_region    complement(26240. .26731)
                      /rpt_family="L1MDb"
     repeat_region    complement(27000. .27314)
                      /rpt_family="AluSx"
     repeat_region    28028. .28099
                      /rpt_family="CT-rich"
     repeat_region    28457. .28714
                      /rpt_family="L2"
     repeat_region    29028. .29059
                      /rpt_family="AT_rich"
     repeat_region    complement(31477. .31543)
                      /rpt_family="L2"
     repeat_region    31875. .31951
                      /rpt_family="MIR3"
     repeat_region    complement(32400. .32584)
                      /rpt_family="MER5B"
     repeat_region    33040. .33325
                      /rpt_family="AluSx"
     repeat_region    33326. .33396
                      /rpt_family="(CAATA)n"
     repeat_region    33593. .33803
                      /rpt_family="L1MB4"
     repeat_region    34265. .34460
                      /rpt_family="MER106A"
     repeat_region    34490. .34529
                      /rpt_family="(A)n"
     repeat_region    complement(34705. .34760)
                      /rpt_family="L2"
     repeat_region    35479. .35677
                      /rpt_family="MER69"
     repeat_region    complement(36643. .36838)
                      /rpt_family="L1ME"
     repeat_region    36865. .38843
                      /rpt_family="L1ME1"
     repeat_region    38849. .39029
                      /rpt_family="MER8"
     repeat_region    complement(39054. .39136)
                      /rpt_family="Tigger2"

  Query Match             42.8%;  Score 27.4;  DB 9;  Length 132288;
  Best Local Similarity   69.8%;  Pred. No. 40;
  Matches   37;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       5 AAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACAT 57
           ||| ||  |  || || ||  ||||||||||    ||||||||| | || ||
Db   45046 AAACAGATCAATGGTATGGAATGAAAAGTTCAGAAACAGACCCATATATATAT 45098
```

Search completed: May 5, 2003, 19:24:44
Job time : 557.297 secs

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:50:42 ; Search time 50.6374 Seconds
(without alignments)
2846.276 Million cell updates/sec

Title: US-09-990-099-23
Perfect score: 64
Sequence: 1 caaaaaaaagtccgctgata.....cccatttttacatcgtagcc 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	64	100.0	815	23	AAS93669	DNA encoding novel
c	2	28.4	44.4	758	22	AAI94502	Human neuroblastom
	3	26.2	40.9	1419	23	AAS71401	DNA encoding novel
	4	25.6	40.0	4313	24	ABQ70923	Listeria monocytog
	5	25.6	40.0	2944528	24	ABA03041	Listeria monocytog
	6	25.4	39.7	8365	20	AAZ20056	Plasmodium falcipa
c	7	25	39.1	700	22	AAH92117	Human inflammatory
c	8	24.8	38.8	2363	23	ABL27140	Drosophila melanog
c	9	24.8	38.8	4001	23	ABL27142	Drosophila melanog
	10	24.6	38.4	1484	22	AAS44601	Human full-length
c	11	24.6	38.4	1484	22	AAS44601	Human full-length
c	12	24.6	38.4	1865	19	AAV41728	DNA encoding the m
c	13	24.6	38.4	1940	19	AAV41732	Codon-optimised RA
c	14	24.6	38.4	2015	7	AAN60107	Sequence encoding
c	15	24.6	38.4	2064	10	AAN92583	Sequence of human
c	16	24.6	38.4	2064	18	AAT66701	Human serum albumi
c	17	24.6	38.4	2251	4	AAN30050	Sequence encoding
c	18	24.6	38.4	2251	20	AAX40195	Human albumin gene
c	19	24.6	38.4	2258	7	AAN60078	Recombinant human
c	20	24.6	38.4	6792	22	AAA54552	Human serum albumi
c	21	24.6	38.4	7934	22	AAS45446	Chemically pretrea
c	22	24.6	38.4	7934	24	AAS61343	Human gene regulat
c	23	24.6	38.4	7934	24	ABK28295	DNA transcription
c	24	24.6	38.4	19012	14	AAQ36952	HSA gene. Homo sa
	25	24.4	38.1	443	19	AAV53480	DNA encoding a Sta
c	26	24.4	38.1	826	24	AAS62774	cDNA sequence #561
c	27	24.4	38.1	3948	23	ABL13622	Drosophila melanog
c	28	24.4	38.1	4923	18	AAV74862	Staphylococcus aur
c	29	24.2	37.8	1043	24	ABK79961	Bacillus clausii g
	30	24.2	37.8	1228	24	ABK77063	Bacillus lichenifo
	31	24.2	37.8	163319	21	AAF22306	Arabidopsis thalia
	32	24	37.5	986	21	AAC34222	Arabidopsis thalia
	33	24	37.5	1317	21	AAC46130	Arabidopsis thalia
	34	24	37.5	1349	21	AAC37629	Arabidopsis thalia
	35	24	37.5	3999	21	AAA71993	A. thaliana FKBP-1
	36	24	37.5	6000	17	AAT06308	Arabidopsis pathog
	37	24	37.5	7000	17	AAT37476	RPP5 downy mildew
c	38	24	37.5	495269	24	ABQ67195	Listeria innocua c
	39	24	37.5	3011208	24	ABQ69245	Listeria innocua D
	40	23.8	37.2	763	20	AAZ15091	Human gene express
	41	23.8	37.2	852	24	ABN68904	Streptococcus poly
	42	23.8	37.2	1371	23	AAS53339	Haemophilus influe
	43	23.8	37.2	2477	23	ABL28098	Drosophila melanog
c	44	23.8	37.2	6583	22	ABA19235	Human nervous syst
	45	23.8	37.2	1830121	17	AAT42063	Haemophilus influe

ALIGNMENTS

RESULT 1
AAS93669
ID AAS93669 standard; cDNA; 815 BP.
XX
AC AAS93669;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29473.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG29482.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 29473; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 815 BP; 210 A; 203 C; 187 G; 215 T; 0 other;

Query Match 100.0%; Score 64; DB 23; Length 815;
Best Local Similarity 100.0%; Pred. No. 4.4e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 100 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 159

Qy 61 AGCC 64
||||
Db 160 AGCC 163

RESULT 2
AAI94502/c
ID AAI94502 standard; cDNA; 758 BP.
XX
AC AAI94502;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 577.
XX
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO200166719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP01629.
XX
PR 07-MAR-2000; 2000JP-0159195.
XX
PA (CHIB-) CHIBA PREFECTURE.
PA (HISM) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX
DR WPI; 2001-565584/63.
XX
PT Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents -
XX
PS Claim 1; Page 464; 2979pp; Japanese.
XX
CC The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
XX
SQ Sequence 758 BP; 193 A; 162 C; 180 G; 194 T; 29 other;

Query Match 44.4%; Score 28.4; DB 22; Length 758;
Best Local Similarity 64.4%; Pred. No. 1.7;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 3 AAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
|| | |||||| | | | |||| | || ||||||| | ||| |||| | |||
Db . 711 AAGGACGGTCCGCNGTTTAANANTGAACAATTAATTTCCAANNCAATTATTACGTTGTA 653

RESULT 3
AAS71401
ID AAS71401 standard; cDNA; 1419 BP.
XX
AC AAS71401;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #7205.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG07214.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 7205; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating

```
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1419 BP; 281 A; 435 C; 403 G; 300 T; 0 other;

  Query Match             40.9%;  Score 26.2;  DB 23;  Length 1419;
  Best Local Similarity   72.3%;  Pred. No. 12;
  Matches   34;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      10 GTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACA 56
           ||| ||| || ||||| || |||||| | |||| | ||||   | ||
Db    1277 GTCTGCTCATGAGGCTAGACAAGTTCCTCTCCATAACCATCCATCCA 1323


RESULT 4
ABQ70923
ID   ABQ70923 standard; DNA; 4313 BP.
XX
AC   ABQ70923;
XX
DT   29-AUG-2002  (first entry)
XX
DE   Listeria monocytogenes 4b contig DNA sequence #865.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria monocytogenes 4b.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators -
XX
PS   Claim 14; SEQ ID 3736; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4313 BP; 1414 A; 694 C; 1022 G; 1183 T; 0 other;

  Query Match             40.0%;  Score 25.6;  DB 24;  Length 4313;
  Best Local Similarity   66.1%;  Pred. No. 24;
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACAT 57
           || ||||| ||||   | |||||||  ||  |  |||| ||||    |||||  ||
Db    3169 AACAAAAATTCCGTGAAAAAGGCTTTGAAGTTCTATTTTCAGATGGCTTTTTCAAT 3224


RESULT 5
ABA03041
ID   ABA03041 standard; DNA; 2944528 BP.
XX
AC   ABA03041;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Listeria monocytogenes EGD-e genome sequence.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW   vitamin B12; bacterial infection; disease; ds.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C,  Frangeul L,  Couve E,  Rusniok C,  Fsihi H,  Dehoux P;
PI   Dussurget O,  Chetouani F,  Nedjari H,  Glaser P,  Kunst F,  Cossart P;
PI   Daniels J,  Goebel W,  Kreft J,  Kuhn M,  Ng E,  Vazquez-Boland JA;
PI   Dominguez-Bernal G,  Garrido-Garcia P,  Tierrez-Martinez A,  Amend A;
PI   Chakraborty T,  Domann E,  Hain T,  Berche P,  Charbit A,  Durant L;
PI   Perez-Diaz J,  Baquero F,  Garcia Del Portillo F,  Gomez-Lopez N;
PI   Maduenio E,  De Pablos B,  Wehland J,  Kaerst U,  Entian K,  Hauf J;
PI   Rose M,  Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides -
XX
PS   Claim 1; SEQ ID No 1; 192pp; French.
XX
CC   The present sequence is the genome sequence of Listeria monocytogenes
CC   EGD-e. This sequence and fragments of this sequence are useful for
CC   selecting probes and primers for detecting genes in L. monocytogenes and
CC   related organisms, and to study genetic polymorphisms and other genomes.
CC   Proteins (ABB47297-ABB50149) expressed from the present sequence are
CC   useful for raising specific antibodies, identification of L.
CC   monocytogenes and related organisms, and for biosynthesis and
CC   biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC   proteins encoded by it are also useful for selecting compounds that
CC   regulate gene expression and cell replication and modulate L.
CC   monocytogenes-related diseases. In addition, this sequence and proteins
CC   encoded by it are useful in pharmaceutical and vaccines compositions for
CC   the treatment or prevention of infections by L. monocytogenes and related
CC   organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

  Query Match             40.0%;  Score 25.6;  DB 24;  Length 2944528;
  Best Local Similarity   66.1%;  Pred. No. 85;
```

```
  Matches   37;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACAT 57
           || ||||| ||||   | |||||||  ||  |  |||| ||||    |||||  ||
Db 1375218 AACAAAAATTCCGTGAAAAAGGCTTTGAAGTTCTATTTTCAGATGGTTTTTTCAAT 1375273


RESULT 6
AAZ20056
ID   AAZ20056 standard; DNA; 8365 BP.
XX
AC   AAZ20056;
XX
DT   21-DEC-1999  (first entry)
XX
DE   Plasmodium falciparum cytoadherence gene clag9 paralogue.
XX
KW   Clag9; paralogue; cytoadherence linked asexual gene; CLAG;
KW   erythrocyte; red blood cell; malaria; infection; therapy; vaccine;
KW   ss.
XX
OS   Plasmodium falciparum.
XX
PN   WO9949048-A1.
XX
PD   30-SEP-1999.
XX
PF   25-MAR-1999;   99WO-AU00213.
XX
PR   25-MAR-1998;   98AU-0002580.
XX
PA   (MENZ-) MENZIES SCHOOL HEALTH RES.
XX
PI   Kemp DJ,  Trenholme KR,  Gardiner DL,  Holt DC,  Cowman AF;
XX
DR   WPI; 1999-591099/50.
XX
PT   New proteins useful for treatment of Plasmodium infections in humans,
PT   especially malaria  -
XX
PS   Disclosure; Page 85-89; 150pp; English.
XX
CC   This is the nucleotide sequence of a clag9 paralogue gene on
CC   chromosome 3 of Plasmodium falciparum.  It is a paralogue of the
CC   newly identified cytoadherence linked asexual gene 9 (clag9) on
CC   chromosome 9 of P. falciparum.  The clag9 gene encodes a protein
CC   (see AAY31945) that facilitates cytoadherence of erythrocytes
CC   parasitised with Plasmodium to other cells.  The identification of
CC   clag genetic species, and the products encoded by them, enables a
CC   range of therapeutic agents to be rationally designed and/or
CC   identified that are useful for the prophylaxis and treatment of
CC   disease conditions caused or exacerbated by infection with
CC   Plasmodium spp., e.g. malaria, especially in humans.
XX
SQ   Sequence 8365 BP; 3570 A; 931 C; 981 G; 2883 T; 0 other;

  Query Match             39.7%;  Score 25.4;  DB 20;  Length 8365;
  Best Local Similarity   64.4%;  Pred. No. 32;
  Matches   38;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       3 AAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||||  ||  ||| ||| |   | |  | ||| | ||||||   |||||| |||  ||
Db    8298 AAAAATTGTATGCTTATATGATGTTATCAATTCTTATCCAGATAAATTTTTTCATAATA 8356


RESULT 7
AAH92117/c
ID   AAH92117 standard; DNA; 700 BP.
XX
AC   AAH92117;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Human inflammatory bowel disease related gene fragment IGR2129a.
XX
KW   Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;
KW   single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;
KW   chromosome 5q31-33; forensic test; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200142511-A2.
XX
PD   14-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US33632.
XX
PR   10-DEC-1999;   99US-0170257.
PR   10-APR-2000; 2000US-0196046.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (ELLI-) ELLIPSIS BIOTHERAPEUTICS CORP.
XX
PI   Daly M,  Hudson TJ,  Lander ES,  Rioux J,  Siminovitch K;
XX
DR   WPI; 2001-367874/38.
XX
PT   Testing for the presence of polymorphisms associated with inflammatory
PT   bowel disease, using a hybridization assay -
XX
PS   Disclosure; Page 123; 463pp; English.
XX
CC   The present invention describes a method for detecting the presence of
CC   polymorphisms associated with inflammatory bowel diseases such as
CC   ulcerative colitis and Crohn's disease. The methods can be used to detect
CC   the presence of genetic polymorphisms associated with inflammatory bowel
CC   disease and correlating their occurrence with disease states. They may be
CC   used in this way for phenotypic correlations, forensics, paternity
CC   testing, medicine and genetic analysis. The present sequence is a gene
CC   containing a polymorphic site described in the exemplification of the
CC   invention.
XX
SQ   Sequence 700 BP; 204 A; 139 C; 165 G; 192 T; 0 other;

  Query Match             39.1%;  Score 25;  DB 22;  Length 700;
  Best Local Similarity   69.4%;  Pred. No. 27;
  Matches   34;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       8 AAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACA 56
           |||||  |  |  || | ||||  | |||||||  |||||  ||||| ||
Db     384 AAGTCACCAAACCAGCCCTGAACTGATCATTTCTAGACCTTTTTTTTCA 336


RESULT 8
ABL27140/c
ID   ABL27140 standard; DNA; 2363 BP.
XX
AC   ABL27140;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 32893.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
```

XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 32893; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2363 BP; 729 A; 471 C; 490 G; 673 T; 0 other;

Query Match 38.8%; Score 24.8; DB 23; Length 2363;
Best Local Similarity 72.7%; Pred. No. 40;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 11 TCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTA 54
|||| |||| | | || |||||||||||| || || ||||
Db 591 TCCGATGATTAAATTCAAAGAGTTCATTTCCACACACAACTTTA 548

RESULT 9
ABL27142/c
ID ABL27142 standard; DNA; 4001 BP.
XX
AC ABL27142;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 32899.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 32899; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4001 BP; 1135 A; 799 C; 940 G; 1127 T; 0 other;

Query Match 38.8%; Score 24.8; DB 23; Length 4001;
Best Local Similarity 72.7%; Pred. No. 45;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 11 TCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTA 54
|||| |||| | | || |||||||||||| || || ||||
Db 3152 TCCGATGATTAAATTCAAAGAGTTCATTTCCACACACAACTTTA 3109

RESULT 10
AAS44601
ID AAS44601 standard; DNA; 1484 BP.
XX
AC AAS44601;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #26.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI; 2001-589862/66.
DR P-PSDB; AAU27701.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
PS Claim 1; SEQ ID No 26; 153pp; English.
XX

CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC contig polynucleotides encoding polypeptides of the invention. The DNA
CC and protein sequences are useful for the treatment, diagnosis and
CC prevention of various types of disorder in a mammalian subject such as a
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1484 BP; 448 A; 297 C; 293 G; 446 T; 0 other;

```
  Query Match           38.4%;  Score 24.6;  DB 22;  Length 1484;
  Best Local Similarity 61.9%;  Pred. No. 43;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||   |  ||||||| |   | | | ||  ||| |||
Db     120 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 179

Qy      62 GCC 64
           | |
Db     180 GGC 182
```

RESULT 11
AAS44601/c
ID AAS44601 standard; DNA; 1484 BP.
XX
AC AAS44601;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #26.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI; 2001-589862/66.
DR P-PSDB; AAU27701.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
PS Claim 1; SEQ ID No 26; 153pp; English.
XX
CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC contig polynucleotides encoding polypeptides of the invention. The DNA
CC and protein sequences are useful for the treatment, diagnosis and
CC prevention of various types of disorder in a mammalian subject such as a
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1484 BP; 448 A; 297 C; 293 G; 446 T; 0 other;

```
  Query Match           38.4%;  Score 24.6;  DB 22;  Length 1484;
  Best Local Similarity 61.9%;  Pred. No. 43;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||   |  ||||||| |   | | | ||  ||| |||
Db    1369 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 1310

Qy      62 GCC 64
           | |
Db    1309 GGC 1307
```

RESULT 12
AAV41728/c
ID AAV41728 standard; DNA; 1865 BP.
XX
AC AAV41728;
XX
DT 20-NOV-1998 (first entry)
XX
DE DNA encoding the mature protein of human serum albumin (HSA).
XX
KW Protein expression; monocotyledon plant cell;
KW glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;
KW ATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;
KW antithrombotic; blood replacement; ss.
XX
OS Homo sapiens.
XX
PN WO9836085-A1.
XX

PD 20-AUG-1998.
XX
PF 13-FEB-1998; 98WO-US03068.
XX
PR 13-FEB-1997; 97US-0038170.
PR 13-FEB-1997; 97US-0037991.
PR 13-FEB-1997; 97US-0038168.
PR 13-FEB-1997; 97US-0038169.
XX
PA (PHYT-) APPLIED PHYTOLOGICS INC.
XX
PI Rodriguez RL, Sutliff TD;
XX
DR WPI; 1998-467179/40.
DR P-PSDB; AAW59841.
XX
PT Expressing mature, glycosylated proteins in monocotyledonous plant
PT cells - from chimeric gene including signal peptide sequence,
PT specifically therapeutic agents and industrial enzymes
XX
PS Disclosure; Pages 32-33; 53pp; English.
XX
CC The present sequence encodes the mature protein of human serum albumin
CC (HSA). The protein is used to exemplify the invention. The
CC specification describes a method for producing mature heterologous
CC protein in monocotyledonous plant cells. The method comprises
CC transforming the cells with a chimeric gene comprising a monocctyledon
CC transcription regulator, inducible either during seed maturation or by
CC adding/removing a small molecule, DNA encoding the heterologous protein,
CC and DNA encoding a signal peptide, with the signal peptide causing
CC secretion of the protein from the cell. Proteins expressed in this
CC manner include mature glycosylated alpha 1-antitrypsin (AAT) with a
CC glycosylation pattern that significantly increases its serum half-life,
CC mature glycosylated antithrombin III (ATIII), mature human serum albumin
CC (HSA) having the native folding pattern as shown by bilirubin-binding
CC characteristics, or mature active subtilisin BPN'. These proteins are
CC useful therapeutically (e.g. AAT for treating emphysema, ATIII as
CC antithrombotic and HSA as blood replacement) or as industrial enzymes
CC (BPN' is used in detergents).
XX
SQ Sequence 1865 BP; 587 A; 369 C; 422 G; 487 T; 0 other;

Query Match 38.4%; Score 24.6; DB 19; Length 1865;
Best Local Similarity 61.9%; Pred. No. 45;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           ||||| ||  | | |||  | ||||   |  |||||||  |   | | | ||  ||| |||
Db    1845 AAAAAGAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 1786

Qy      62 GCC 64
           | |
Db    1785 GGC 1783
```

RESULT 13
AAV41732/c
ID AAV41732 standard; DNA; 1940 BP.
XX
AC AAV41732;
XX
DT 20-NOV-1998 (first entry)
XX
DE Codon-optimised RAmy3D signal fused to DNA encoding mature HSA.
XX
KW Protein expression; monocotyledon plant cell;
KW glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;
KW ATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;
KW antithrombotic; blood replacement; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX

FH Key Location/Qualifiers
FT misc_feature 1..75
FT /*tag= a
FT /note= "codon-optimised RAmy3D signal sequence"
FT misc_feature 76..1940
FT /*tag= b
FT /note= "encodes mature HSA"
XX
PN WO9836085-A1.
XX
PD 20-AUG-1998.
XX
PF 13-FEB-1998; 98WO-US03068.
XX
PR 13-FEB-1997; 97US-0038170.
PR 13-FEB-1997; 97US-0037991.
PR 13-FEB-1997; 97US-0038168.
PR 13-FEB-1997; 97US-0038169.
XX
PA (PHYT-) APPLIED PHYTOLOGICS INC.
XX
PI Rodriguez RL, Sutliff TD;
XX
DR WPI; 1998-467179/40.
XX
PT Expressing mature, glycosylated proteins in monocotyledonous plant
PT cells - from chimeric gene including signal peptide sequence,
PT specifically therapeutic agents and industrial enzymes
XX
PS Disclosure; Pages 34 iv-v; 53pp; English.
XX
CC The present sequence encodes a fusion protein of codon-optimised RAmy3D
CC signal sequence/mature human serum albumin (HSA). The protein is used
CC to exemplify the invention. The specification describes a method for
CC producing mature heterologous protein in monocotyledonous plant cells.
CC The method comprises transforming the cells with a chimeric gene
CC comprising a monocotyledon transcription regulator, inducible either
CC during seed maturation or by adding/removing a small molecule, DNA
CC encoding the heterologous protein, and DNA encoding a signal peptide,
CC with the signal peptide causing secretion of the protein from the cell.
CC Proteins expressed in this manner include mature glycosylated alpha
CC 1-antitrypsin (AAT) with a glycosylation pattern that significantly
CC increases its serum half-life, mature glycosylated antithrombin III
CC (ATIII), mature human serum albumin (HSA) having the native folding
CC pattern as shown by bilirubin-binding characteristics, or mature active
CC subtilisin BPN'. These proteins are useful therapeutically (e.g. AAT for
CC treating emphysema, ATIII as antithrombotic and HSA as blood replacement)
CC or as industrial enzymes (BPN' is used in detergents).
XX
SQ Sequence 1940 BP; 598 A; 401 C; 438 G; 503 T; 0 other;

Query Match 38.4%; Score 24.6; DB 19; Length 1940;
Best Local Similarity 61.9%; Pred. No. 45;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           ||||| ||  | | |||  | ||||   |  |||||||  |   | | | ||  ||| |||
Db    1920 AAAAAGAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 1861

Qy      62 GCC 64
           | |
Db    1860 GGC 1858
```

RESULT 14
AAN60107/c
ID AAN60107 standard; cDNA; 2015 BP.
XX
AC AAN60107;
XX
DT 29-JUN-1991 (first entry)
XX
DE Sequence encoding prepro human serum albumin (HSA).

```
XX
KW   Shock therapy; hypoproteinaemia; erythroblastosis foetalis;
KW   burn therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             32..85
FT                   /*tag=  a
FT                   /product= pre sequence
FT   CDS             86..103
FT                   /*tag=  b
FT                   /product= pro sequence
FT   mat_peptide     104..1861
FT                   /*tag=  c
XX
PN   EP206733-A.
XX
PD   30-DEC-1986.
XX
PF   17-JUN-1986;    86EP-0304656.
XX
PR   17-JUN-1985;    85US-0745524.
PR   17-JUN-1986;    86JP-0141206.
PR   19-DEC-1986;    86JP-0303613.
XX
PA   (GENE-) GENEX CORP.
PA   (MITQ ) MITSUBISHI DENKI KK.
XX
PI   Burns AL;
XX
DR   WPI; 1986-341032/52.
DR   P-PSDB; AAP60092.
XX
PT   New synthetic human serum albumin gene - useful in plasmids for
PT   transforming microorganisms
XX
PS   Disclosure; Fig 2; 44pp; English.
XX
CC   An E.coli HB101 culture transformed with the plasmid (pGX401)
CC   containing the prepro-HSA gene (AAN60107) has been deposited with the
CC   U.S. Department of Agriculture Northern Regional Research Laboratory
CC   in Peoria, Illinois, as NRRL No. B-15784. pGX401 is specifically
CC   claimed. It is believed that the AA sequence set forth in  AAP60092
CC   represents a genomic HSA allele that is widespread in the human
CC   population. AAN60106 includes all variations of the HSA gene permitted
CC   by the genetic code.
XX
SQ   Sequence 2015 BP; 612 A; 406 C; 447 G; 550 T; 0 other;

  Query Match             38.4%;  Score 24.6;  DB 7;  Length 2015;
  Best Local Similarity   61.9%;  Pred. No. 46;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           ||||| ||  | | |||  | ||||   |  |||||||| |   | | | ||  ||| |||
Db    1947 AAAAAGAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 1888

Qy      62 GCC 64
           | |
Db    1887 GGC 1885


RESULT 15
AAN92583/c
ID   AAN92583 standard; DNA; 2064 BP.
XX
AC   AAN92583;
XX
DT   13-MAR-1992  (first entry)
XX
DE   Sequence of human serum albumin 13 (HSA 13).
XX
KW   Albumin; yeast expression vector; methylotrophic yeast; ss.
XX
OS   Homo sapiens.
XX
PN   EP344459-A.
XX
PD   06-DEC-1989.
XX
PF   25-APR-1989;    89EP-0107459.
XX
PR   25-APR-1988;    88US-0186420.
XX
PA   (PHIP ) PHILLIPS PETROLEUM CO.
XX
PI   Marashi F,  Fuke M,  Mcombie WR,  Barr RD,  Sreekishna K;
PI   Wong R;
XX
DR   WPI; 1989-357919/49.
XX
PT   HSA prodn. in yeast - comprises methylotrophic yeast transformed
PT   with vectors encoding HSA
XX
PS   Claim 24; Table 1, pages 4-5; 28pp; English.
XX
CC   The inventors claim the sequence in AAN92583, and vectors contg. a
CC   structural HSA gene, a plasmid for expression in yeast and
CC   methylotrophic yeast transfromed by the vector. The vector
CC   pHSA113 and Pichia pastoris 95115/pHSA113 are specifically claimed.
XX
SQ   Sequence 2064 BP; 649 A; 406 C; 448 G; 561 T; 0 other;

  Query Match             38.4%;  Score 24.6;  DB 10;  Length 2064;
  Best Local Similarity   61.9%;  Pred. No. 46;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           ||| ||||  | | |||  | ||||   |  |||||||| |   | | | ||  ||| |||
Db    1940 AAAGAAAACACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 1881

Qy      62 GCC 64
           | |
Db    1880 GGC 1878


Search completed: May  5, 2003, 18:05:50
Job time : 440.637 secs
```

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:51:17 ; Search time 11.6044 Seconds
(without alignments)
1691.369 Million cell updates/sec

Title: US-09-990-099-23
Perfect score: 64
Sequence: 1 caaaaaaaagtccgctgata.....cccatttttacatcgtagcc 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	24.6	38.4	231	1	US-07-691-079C-8	Sequence 8, Appli
c	2	24.6	38.4	231	1	US-08-433-037-18	Sequence 18, Appl
c	3	24.6	38.4	19011	1	US-08-310-356-36	Sequence 36, Appl
c	4	24.6	38.4	19557	5	PCT-US92-06300-1	Sequence 1, Appli
	5	24.4	38.1	443	4	US-08-936-165A-180	Sequence 180, App
c	6	23.4	36.6	4451	4	US-09-303-064-45	Sequence 45, Appl
c	7	23.4	36.6	4451	4	US-09-086-503-45	Sequence 45, Appl
c	8	23.4	36.6	4481	4	US-08-867-611-1	Sequence 1, Appli
c	9	23.4	36.6	4481	5	PCT-US92-06965A-6	Sequence 6, Appli
c	10	23.4	36.6	4775	4	US-09-303-064-37	Sequence 37, Appl
c	11	23.4	36.6	4775	4	US-09-086-503-37	Sequence 37, Appl
c	12	23.4	36.6	4910	4	US-09-303-064-40	Sequence 40, Appl
c	13	23.4	36.6	4910	4	US-09-086-503-40	Sequence 40, Appl
c	14	23.4	36.6	5258	4	US-09-303-064-48	Sequence 48, Appl
c	15	23.4	36.6	5258	4	US-09-086-503-48	Sequence 48, Appl
c	16	23.4	36.6	5600	4	US-08-867-611-3	Sequence 3, Appli
c	17	23.4	36.6	5600	5	PCT-US92-06965A-8	Sequence 8, Appli
c	18	23.2	36.2	3212	4	US-08-855-910-1	Sequence 1, Appli
c	19	23.2	36.2	10614	1	US-08-135-511-35	Sequence 35, Appl
c	20	23.2	36.2	10614	1	US-08-187-453-35	Sequence 35, Appl
c	21	23	35.9	2058	4	US-09-560-639-4	Sequence 4, Appli
c	22	23	35.9	2504	2	US-08-946-412-1	Sequence 1, Appli
c	23	22.6	35.3	2321	4	US-09-198-839-2	Sequence 2, Appli
c	24	22.6	35.3	2321	4	US-09-562-834-2	Sequence 2, Appli
c	25	22.6	35.3	2337	4	US-09-198-839-3	Sequence 3, Appli
c	26	22.6	35.3	2337	4	US-09-562-834-3	Sequence 3, Appli
	27	22.4	35.0	8651	4	US-08-961-527-181	Sequence 181, App
	28	22.2	34.7	75	1	US-07-770-100C-2	Sequence 2, Appli
	29	22.2	34.7	1929	4	US-08-750-180-1	Sequence 1, Appli
	30	22.2	34.7	43804	4	US-09-171-461-1	Sequence 1, Appli
	31	22	34.4	3155	2	US-08-591-629-7	Sequence 7, Appli
	32	21.8	34.1	1160	4	US-08-858-207A-209	Sequence 209, App
c	33	21.8	34.1	2004	1	US-08-471-033-6	Sequence 6, Appli
c	34	21.8	34.1	2004	2	US-08-471-044-6	Sequence 6, Appli
c	35	21.8	34.1	2004	2	US-08-463-483A-6	Sequence 6, Appli
c	36	21.8	34.1	2004	2	US-08-471-046A-6	Sequence 6, Appli
c	37	21.8	34.1	2004	2	US-08-470-566B-6	Sequence 6, Appli
c	38	21.8	34.1	2004	2	US-08-469-334-6	Sequence 6, Appli
c	39	21.8	34.1	2004	3	US-09-300-529-6	Sequence 6, Appli
c	40	21.8	34.1	2408	2	US-08-870-827-5	Sequence 5, Appli
c	41	21.8	34.1	2408	4	US-09-317-179-5	Sequence 5, Appli
c	42	21.8	34.1	2645	4	US-08-960-780-31	Sequence 31, Appl
c	43	21.8	34.1	2645	4	US-09-073-898-31	Sequence 31, Appl
c	44	21.8	34.1	2645	4	US-09-371-913A-1	Sequence 1, Appli
c	45	21.8	34.1	2655	1	US-08-471-033-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-07-691-079C-8/c
; Sequence 8, Application US/07691079C
; Patent No. 5330901
;  GENERAL INFORMATION:
;    APPLICANT:  William D. Prevatt et al.
;    TITLE OF INVENTION:  Expression of Human Serum Albumin in
;    TITLE OF INVENTION:  Pichia pastoris
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RICHMOND, PHILLIPS, HITCHCOCK & UMPHLETT
;      STREET:  P.O. Box 2443
;      CITY:  Bartlesville
;      STATE:  OK.
;      COUNTRY:  USA
;      ZIP:  74005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM PC
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Display Write 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/691,079C
;      FILING DATE:  19910426
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hal Brent Woodrow
;      REGISTRATION NUMBER:  32,501
;      REFERENCE/DOCKET NUMBER:  32819US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-918-661-0624
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  231bp
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Linker Oligonucleotide
US-07-691-079C-8

  Query Match             38.4%;  Score 24.6;  DB 1;  Length 231;
  Best Local Similarity   61.9%;  Pred. No. 3.4;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           ||| ||||  | | |||  | ||||   |  |||||||  |   | | | ||  ||| |||
Db     102 AAAGAAAACACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 43

Qy      62 GCC 64
           | |
```

```
Db      42 GGC 40


RESULT 2
US-08-433-037-18/c
; Sequence 18, Application US/08433037
; Patent No. 5707828
;  GENERAL INFORMATION:
;    APPLICANT:  Sreekrishna, Kotikanyadan
;    APPLICANT:  Barr, Kathryn A.
;    APPLICANT:  Brierley, Russell A.
;    APPLICANT:  Thill, Gregory P.
;    APPLICANT:  Tschopp, Juerg F.
;    TITLE OF INVENTION:  EXPRESSION OF HUMAN SERUM ALBUMIN IN
;    TITLE OF INVENTION:  PICHIA PASTORIS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Scully, Scott, Murphy & Presser
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  11530-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,037
;      FILING DATE:  03-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  9108Z
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (516) 742-4343
;      TELEFAX:  (516) 742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  231 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-433-037-18

  Query Match             38.4%;  Score 24.6;  DB 1;  Length 231;
  Best Local Similarity   61.9%;  Pred. No. 3.4;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           ||| ||||  | | |||  | ||||   |  ||||||| |   | | | ||  ||| |||
Db     102 AAAGAAAACACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 43

Qy      62 GCC 64
           | |
Db      42 GGC 40


RESULT 3
US-08-310-356-36/c
; Sequence 36, Application US/08310356
; Patent No. 5648243
;  GENERAL INFORMATION:
;    APPLICANT:  Hurwitz, David R
;    APPLICANT:  Nathan, Margret
;    APPLICANT:  Shani, Moshe
;    TITLE OF INVENTION:  Transgenic Protein Production
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rhone-Poulenc Rorer Legal Department
;      STREET:  500 Arcola Road
;      CITY:  Collegeville
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19426
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh System 7.0
;      SOFTWARE:  Microsoft Word Version 5.0 (PatentIn)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/310,356
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/737,853
;      FILING DATE:  31-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goodman, Rosanne
;      REGISTRATION NUMBER:  32,534
;      REFERENCE/DOCKET NUMBER:  A0856
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 454-3817
;      TELEFAX:  (215) 454-3808
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19011 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1737..1854
;      OTHER INFORMATION:  /number= 1
;      OTHER INFORMATION:  /citation= ([1])
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1855..2563
;      OTHER INFORMATION:  /number= 1
;      OTHER INFORMATION:  /label= intron1
;      OTHER INFORMATION:  /citation= ([1])
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2564..2621
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /citation= ([1])
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  2622..4075
;      OTHER INFORMATION:  /number= 2
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4076..4208
;      OTHER INFORMATION:  /number= 3
;      OTHER INFORMATION:  /citation= ([1])
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  4209..6040
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  6041..6252
;      OTHER INFORMATION:  /number= 4
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  6253..6801
;      OTHER INFORMATION:  /number= 4
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  6802..6934
;      OTHER INFORMATION:  /number= 5
```

; FEATURE:
; NAME/KEY: intron
; LOCATION: 6935..7758
; OTHER INFORMATION: /number= 5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 7759..7856
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: intron
; LOCATION: 7857..9443
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 9444..9573
; OTHER INFORMATION: /number= 7
; FEATURE:
; NAME/KEY: intron
; LOCATION: 9574..10866
; OTHER INFORMATION: /number= 7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10867..11081
; OTHER INFORMATION: /number= 8
; FEATURE:
; NAME/KEY: intron
; LOCATION: 11082..12480
; OTHER INFORMATION: /number= 8
; FEATURE:
; NAME/KEY: exon
; LOCATION: 12481..12613
; OTHER INFORMATION: /number= 9
; FEATURE:
; NAME/KEY: intron
; LOCATION: 12614..13701
; OTHER INFORMATION: /number= 9
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13702..13799
; OTHER INFORMATION: /number= 10
; FEATURE:
; NAME/KEY: intron
; LOCATION: 13800..14976
; OTHER INFORMATION: /number= 10
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14977..15115
; OTHER INFORMATION: /number= 11
; FEATURE:
; NAME/KEY: intron
; LOCATION: 15116..15533
; OTHER INFORMATION: /number= 11
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15534..15757
; OTHER INFORMATION: /number= 12
; FEATURE:
; NAME/KEY: intron
; LOCATION: 15758..16949
; OTHER INFORMATION: /number= 12
; FEATURE:
; NAME/KEY: exon
; LOCATION: 16950..17082
; OTHER INFORMATION: /number= 13
; FEATURE:
; NAME/KEY: intron
; LOCATION: 17083..17696
; OTHER INFORMATION: /number= 13
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17697..17764
; OTHER INFORMATION: /number= 14
; FEATURE:
; NAME/KEY: intron
; LOCATION: 17765..18534
; OTHER INFORMATION: /number= 14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1776..1854, 2564..2621, 4076..4208,
; LOCATION: 6041..6252, 6802..6934, 7759..7856, 9444..9573,
; LOCATION: 10867..11081, 12481..12613, 13702..13799,
; LOCATION: 14977..15115, 15534..15757, 16950..17082,
; LOCATION: 17697..17741)
; OTHER INFORMATION: /product= "human serum albumin"
; OTHER INFORMATION: /citation= ([1])
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18535..18697
; OTHER INFORMATION: /number= 15
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 17742..18697
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1737..1775
; PUBLICATION INFORMATION:
; AUTHORS: Minghetti, P P
; AUTHORS: Ruffner, D E
; AUTHORS: Kuang, W-J
; AUTHORS: Dennison, O E
; AUTHORS: Hawkins, J W
; AUTHORS: Beattie, W G
; AUTHORS: Dugaiczyk, A
; TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN
; TITLE: GENE IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN
; TITLE: q11-22 OF CHROMOSOME 4
; JOURNAL: J. Biol. Chem.
; VOLUME: 261
; PAGES: 6747-6757
; DATE: 1986
; RELEVANT RESIDUES IN SEQ ID NO: 36: FROM 1 TO 19011
US-08-310-356-36

Query Match 38.4%; Score 24.6; DB 1; Length 19011;
Best Local Similarity 61.9%; Pred. No. 9.5;
Matches 39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||   |  |||||||  |   | | | ||  ||| |||
Db   18596 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 18537

Qy      62 GCC 64
           | |
Db   18536 GGC 18534
```

RESULT 4
PCT-US92-06300-1/c
; Sequence 1, Application PC/TUS9206300
; GENERAL INFORMATION:
; APPLICANT: Hurwitz, David R
; APPLICANT: Nathan, Margret
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: Transgenic Protein Production
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer, Inc.
; STREET: 500 Virginia Ave., Bldg. 3A
; CITY: Ft. Washington
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19034
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06300
;      FILING DATE:  19920730
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goodman, Rosanne
;      REGISTRATION NUMBER:  52,534
;      REFERENCE/DOCKET NUMBER:  A0856-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 962-4130
;      TELEFAX:  (215) 962-4107
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19557 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    PUBLICATION INFORMATION:
;      AUTHORS:  Minghetti, P P
;      AUTHORS:  Ruffner, D E
;      AUTHORS:  Kuang, W.-J.
;      AUTHORS:  Dennison, O E
;      AUTHORS:  Hawkins, J W
;      AUTHORS:  Beattie, W G
;      AUTHORS:  Dugaiczyk, A
;      TITLE:  MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE
;      TITLE:  IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN Q11-22
;      TITLE:  OF CHROMOSOME 4
;      JOURNAL:  J. Biol. Chem.
;      VOLUME:  261
;      PAGES:  6747-6757
;      DATE:  1986
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 19002
PCT-US92-06300-1

  Query Match             38.4%;  Score 24.6;  DB 5;  Length 19557;
  Best Local Similarity   61.9%;  Pred. No. 9.6;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||   |  |||||||  |   | | | ||  ||| |||
Db   19142 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 19083

Qy      62 GCC 64
           | |
Db   19082 GGC 19080


RESULT 5
US-08-936-165A-180
; Sequence 180, Application US/08936165A
; Patent No. 6348582
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Burnham, Martin
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Lonetto, Michael
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Pratt, Julie
;    APPLICANT:  Reichard, Richard
;    APPLICANT:  Rosenberg, Martin
;    APPLICANT:  Ward, Judith
;    TITLE OF INVENTION:  No. 6348582el Prokaryotic Polynucleotides,
;    TITLE OF INVENTION:  Polypeptides and Their Uses
;    NUMBER OF SEQUENCES:  534
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,165A
;      FILING DATE:  24-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,032
;      FILING DATE:  24-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  180:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  443 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-936-165A-180

  Query Match             38.1%;  Score 24.4;  DB 4;  Length 443;
  Best Local Similarity   73.8%;  Pred. No. 4.7;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      16 TGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACAT 57
           |||| ||| || ||| |||||||| || || ||| |  | ||
Db     200 TGATTAGGTTTCAAAGGTTCATTTTCAAACACATCTAAATAT 241


RESULT 6
US-09-303-064-45/c
; Sequence 45, Application US/09303064
; Patent No. 6221619
; GENERAL INFORMATION:
;  APPLICANT: MAINE, Gregory T.
;  APPLICANT: HUNT, Jeffery C.
;  APPLICANT: BROJANAC, Susan
;  APPLICANT: JYH-TSING SHEU, Michael
;  APPLICANT: CHOVAN, Linda E.
;  APPLICANT: TYNER, Joan D.
;  APPLICANT: HOWARD, Lawrence V.
;  APPLICANT: PARMLEY, Stephen F.
;  APPLICANT: REMINGTON, Jack S.
;  APPLICANT: ARAUJO, Fausto
;  APPLICANT: SUZUKI, Yashuhiro
;  APPLICANT: LI, Shuli
;  TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF
;  FILE REFERENCE: 6361.US.P1
;  CURRENT APPLICATION NUMBER: US/09/303,064
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 09/086,503
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 45
;   LENGTH: 4451
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-303-064-45
```

```
  Query Match             36.6%;  Score 23.4;  DB 4;  Length 4451;
  Best Local Similarity   73.2%;  Pred. No. 18;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
           ||||||||| ||||| || ||||  |  ||| |  || |||
Db    1686 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 1646


RESULT 7
US-09-086-503-45/c
; Sequence 45, Application US/09086503A
; Patent No. 6329157
; GENERAL INFORMATION:
;  APPLICANT: MAINE, Gregory T.
;  APPLICANT: HUNT, Jeffery C.
;  APPLICANT: BROJANAC, Susan
;  APPLICANT: JYH-TSING SHEU, Michael
;  APPLICANT: CHOVAN, Linda E.
;  APPLICANT: TYNER, Joan D.
;  APPLICANT: HOWARD, Lawrence V.
;  TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
;  FILE REFERENCE: 6361.US.01
;  CURRENT APPLICATION NUMBER: US/09/086,503A
;  CURRENT FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 45
;   LENGTH: 4451
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-086-503-45

  Query Match             36.6%;  Score 23.4;  DB 4;  Length 4451;
  Best Local Similarity   73.2%;  Pred. No. 18;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
           ||||||||| ||||| || ||||  |  ||| |  || |||
Db    1686 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 1646


RESULT 8
US-08-867-611-1/c
; Sequence 1, Application US/08867611
; Patent No. 6172189
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, SUSHIL G
;    APPLICANT:  DESAI, SURESH M
;    APPLICANT:  CASEY, JAMES M
;    APPLICANT:  DAILEY, STEPHEN H
;    APPLICANT:  DAWSON, GEORGE J
;    APPLICANT:  GUTIERREZ, ROBIN A
;    APPLICANT:  LESNIEWSKI, RICHARD R
;    APPLICANT:  STEWART, JAMES L
;    APPLICANT:  RUPPRECHT, KEVIN R
;    TITLE OF INVENTION:  HEPATITIS C ASSAY UTILIZING RECOMBINANT
;    TITLE OF INVENTION:  ANTIGENS
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD, CHAD377/AP6D2
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,611
;      FILING DATE:  02-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,757
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/179,896
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/572,822
;      FILING DATE:  24-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/614,069
;      FILING DATE:  07-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/748,561
;      FILING DATE:  21-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/748,565
;      FILING DATE:  21-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/748,566
;      FILING DATE:  21-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  4834.US.P6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4481 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  130..1317
US-08-867-611-1

  Query Match             36.6%;  Score 23.4;  DB 4;  Length 4481;
  Best Local Similarity   73.2%;  Pred. No. 18;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
           ||||||||| ||||| || ||||  |  ||| |  || |||
Db    1716 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 1676


RESULT 9
PCT-US92-06965A-6/c
; Sequence 6, Application PC/TUS9206965A
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, S.
;    APPLICANT:  DESAI, S.
;    APPLICANT:  DAILEY, S.
;    TITLE OF INVENTION:  HCV SYNTHETIC PEPTIDE FROM NS1 REGION
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  U.S.
;      ZIP:  60065-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06965A
```

```
;      FILING DATE:  19920821
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  4834PC.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4481 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  130..1317
PCT-US92-06965A-6

  Query Match             36.6%;  Score 23.4;  DB 5;  Length 4481;
  Best Local Similarity   73.2%;  Pred. No. 18;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
          ||||||||| ||||| || |||| | ||| | || |||
Db   1716 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 1676


RESULT 10
US-09-303-064-37/c
; Sequence 37, Application US/09303064
; Patent No. 6221619
; GENERAL INFORMATION:
;  APPLICANT: MAINE, Gregory T.
;  APPLICANT: HUNT, Jeffery C.
;  APPLICANT: BROJANAC, Susan
;  APPLICANT: JYH-TSING SHEU, Michael
;  APPLICANT: CHOVAN, Linda E.
;  APPLICANT: TYNER, Joan D.
;  APPLICANT: HOWARD, Lawrence V.
;  APPLICANT: PARMLEY, Stephen F.
;  APPLICANT: REMINGTON, Jack S.
;  APPLICANT: ARAUJO, Fausto
;  APPLICANT: SUZUKI, Yashuhiro
;  APPLICANT: LI, Shuli
;  TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF
;  FILE REFERENCE: 6361.US.P1
;  CURRENT APPLICATION NUMBER: US/09/303,064
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 09/086,503
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37
;   LENGTH: 4775
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-303-064-37

  Query Match             36.6%;  Score 23.4;  DB 4;  Length 4775;
  Best Local Similarity   73.2%;  Pred. No. 18;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
          ||||||||| ||||| || |||| | ||| | || |||
Db   2010 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 1970


RESULT 11
US-09-086-503-37/c
; Sequence 37, Application US/09086503A
; Patent No. 6329157
; GENERAL INFORMATION:
;  APPLICANT: MAINE, Gregory T.
;  APPLICANT: HUNT, Jeffery C.
;  APPLICANT: BROJANAC, Susan
;  APPLICANT: JYH-TSING SHEU, Michael
;  APPLICANT: CHOVAN, Linda E.
;  APPLICANT: TYNER, Joan D.
;  APPLICANT: HOWARD, Lawrence V.
;  TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
;  FILE REFERENCE: 6361.US.01
;  CURRENT APPLICATION NUMBER: US/09/086,503A
;  CURRENT FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37
;   LENGTH: 4775
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-086-503-37

  Query Match             36.6%;  Score 23.4;  DB 4;  Length 4775;
  Best Local Similarity   73.2%;  Pred. No. 18;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
          ||||||||| ||||| || |||| | ||| | || |||
Db   2010 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 1970


RESULT 12
US-09-303-064-40/c
; Sequence 40, Application US/09303064
; Patent No. 6221619
; GENERAL INFORMATION:
;  APPLICANT: MAINE, Gregory T.
;  APPLICANT: HUNT, Jeffery C.
;  APPLICANT: BROJANAC, Susan
;  APPLICANT: JYH-TSING SHEU, Michael
;  APPLICANT: CHOVAN, Linda E.
;  APPLICANT: TYNER, Joan D.
;  APPLICANT: HOWARD, Lawrence V.
;  APPLICANT: PARMLEY, Stephen F.
;  APPLICANT: REMINGTON, Jack S.
;  APPLICANT: ARAUJO, Fausto
;  APPLICANT: SUZUKI, Yashuhiro
;  APPLICANT: LI, Shuli
;  TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF
;  FILE REFERENCE: 6361.US.P1
;  CURRENT APPLICATION NUMBER: US/09/303,064
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 09/086,503
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 40
;   LENGTH: 4910
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-303-064-40

  Query Match             36.6%;  Score 23.4;  DB 4;  Length 4910;
  Best Local Similarity   73.2%;  Pred. No. 19;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
          ||||||||| ||||| || |||| | ||| | || |||
Db   2145 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 2105


RESULT 13
US-09-086-503-40/c
; Sequence 40, Application US/09086503A
```

```
; Patent No. 6329157
; GENERAL INFORMATION:
;  APPLICANT: MAINE, Gregory T.
;  APPLICANT: HUNT, Jeffery C.
;  APPLICANT: BROJANAC, Susan
;  APPLICANT: JYH-TSING SHEU, Michael
;  APPLICANT: CHOVAN, Linda E.
;  APPLICANT: TYNER, Joan D.
;  APPLICANT: HOWARD, Lawrence V.
;  TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
;  FILE REFERENCE: 6361.US.01
;  CURRENT APPLICATION NUMBER: US/09/086,503A
;  CURRENT FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 40
;   LENGTH: 4910
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-086-503-40

  Query Match             36.6%;  Score 23.4;  DB 4;  Length 4910;
  Best Local Similarity   73.2%;  Pred. No. 19;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
           |||||||||| ||||| || ||||  |  ||| |  || |||
Db    2145 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 2105


RESULT 14
US-09-303-064-48/c
; Sequence 48, Application US/09303064
; Patent No. 6221619
; GENERAL INFORMATION:
;  APPLICANT: MAINE, Gregory T.
;  APPLICANT: HUNT, Jeffery C.
;  APPLICANT: BROJANAC, Susan
;  APPLICANT: JYH-TSING SHEU, Michael
;  APPLICANT: CHOVAN, Linda E.
;  APPLICANT: TYNER, Joan D.
;  APPLICANT: HOWARD, Lawrence V.
;  APPLICANT: PARMLEY, Stephen F.
;  APPLICANT: REMINGTON, Jack S.
;  APPLICANT: ARAUJO, Fausto
;  APPLICANT: SUZUKI, Yashuhiro
;  APPLICANT: LI, Shuli
;  TITLE OF INVENTION: ANTIGEN COCKTAILS, P35 AND USES THEREOF
;  FILE REFERENCE: 6361.US.P1
;  CURRENT APPLICATION NUMBER: US/09/303,064
;  CURRENT FILING DATE: 1999-04-30
;  EARLIER APPLICATION NUMBER: 09/086,503
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 48
;   LENGTH: 5258
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-303-064-48

  Query Match             36.6%;  Score 23.4;  DB 4;  Length 5258;
  Best Local Similarity   73.2%;  Pred. No. 19;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
           |||||||||| ||||| || ||||  |  ||| |  || |||
Db    2493 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 2453


RESULT 15
US-09-086-503-48/c
; Sequence 48, Application US/09086503A
; Patent No. 6329157
; GENERAL INFORMATION:
;  APPLICANT: MAINE, Gregory T.
;  APPLICANT: HUNT, Jeffery C.
;  APPLICANT: BROJANAC, Susan
;  APPLICANT: JYH-TSING SHEU, Michael
;  APPLICANT: CHOVAN, Linda E.
;  APPLICANT: TYNER, Joan D.
;  APPLICANT: HOWARD, Lawrence V.
;  TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
;  FILE REFERENCE: 6361.US.01
;  CURRENT APPLICATION NUMBER: US/09/086,503A
;  CURRENT FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 48
;   LENGTH: 5258
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-086-503-48

  Query Match             36.6%;  Score 23.4;  DB 4;  Length 5258;
  Best Local Similarity   73.2%;  Pred. No. 19;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
           |||||||||| ||||| || ||||  |  ||| |  || |||
Db    2493 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 2453


Search completed: May  5, 2003, 18:00:52
Job time : 15.6044 secs
```

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:52:42 ; Search time 29.3626 Seconds
(without alignments)
2575.509 Million cell updates/sec

Title: US-09-990-099-23
Perfect score: 64
Sequence: 1 caaaaaaaagtccgctgata.....cccatttttacatcgtagcc 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	26.8	41.9	445	10	US-09-783-590-762	Sequence 762, App
c	2	24.6	38.4	362	9	US-09-918-995-30115	Sequence 30115, A
c	3	24.6	38.4	394	9	US-09-918-995-7625	Sequence 7625, Ap
c	4	24.6	38.4	425	9	US-09-918-995-8119	Sequence 8119, Ap
c	5	24.6	38.4	468	9	US-09-918-995-31162	Sequence 31162, A
	6	24.4	38.1	443	10	US-09-939-980-180	Sequence 180, App
c	7	24.4	38.1	826	10	US-09-822-830A-561	Sequence 561, App
	8	24.4	38.1	1245	9	US-09-938-842A-3526	Sequence 3526, Ap
c	9	24.4	38.1	4923	7	US-08-781-986A-551	Sequence 551, App
c	10	24.2	37.8	1043	10	US-09-974-300-7252	Sequence 7252, Ap
	11	24.2	37.8	1228	10	US-09-974-300-4354	Sequence 4354, Ap
c	12	23.8	37.2	380	10	US-09-878-574-796	Sequence 796, App
	13	23.8	37.2	1371	10	US-09-815-242-6976	Sequence 6976, Ap
c	14	23.6	36.9	173808	12	US-10-003-806-10	Sequence 10, Appl
c	15	23.4	36.6	4451	10	US-09-896-852-45	Sequence 45, Appl
c	16	23.4	36.6	4775	10	US-09-896-852-37	Sequence 37, Appl
c	17	23.4	36.6	4910	10	US-09-896-852-40	Sequence 40, Appl
c	18	23.4	36.6	5258	10	US-09-896-852-48	Sequence 48, Appl
c	19	23.2	36.2	337	9	US-09-796-692-8555	Sequence 8555, Ap
c	20	23.2	36.2	337	9	US-10-040-862-8555	Sequence 8555, Ap
c	21	23.2	36.2	456	9	US-09-918-995-9303	Sequence 9303, Ap
	22	23.2	36.2	1161	7	US-08-781-986A-497	Sequence 497, App
c	23	23.2	36.2	2000	9	US-09-938-842A-4923	Sequence 4923, Ap
c	24	23.2	36.2	2742	10	US-09-815-242-3941	Sequence 3941, Ap
c	25	23.2	36.2	2802	10	US-09-815-242-6546	Sequence 6546, Ap
c	26	23.2	36.2	3779	10	US-09-925-300-712	Sequence 712, App
	27	23.2	36.2	32768	10	US-09-070-927A-128	Sequence 128, App
c	28	23.2	36.2	1503841	9	US-09-946-807-1	Sequence 1, Appli
c	29	23.2	36.2	1503841	10	US-09-795-668-1	Sequence 1, Appli
c	30	23.2	36.2	1503841	10	US-09-795-686-1	Sequence 1, Appli
	31	23	35.9	343	10	US-09-867-701-10855	Sequence 10855, A
c	32	23	35.9	449	10	US-09-770-444-710	Sequence 710, App
	33	23	35.9	498	9	US-09-736-457-312	Sequence 312, App
	34	23	35.9	498	9	US-09-902-941-312	Sequence 312, App
	35	23	35.9	498	9	US-09-849-626-312	Sequence 312, App
	36	23	35.9	498	9	US-10-017-754-312	Sequence 312, App
c	37	23	35.9	571	9	US-10-060-036-1993	Sequence 1993, Ap
	38	23	35.9	711	10	US-09-974-300-4956	Sequence 4956, Ap
	39	23	35.9	780	9	US-09-938-842A-3893	Sequence 3893, Ap
	40	23	35.9	789	10	US-09-974-300-4955	Sequence 4955, Ap
c	41	23	35.9	2058	10	US-09-899-980A-4	Sequence 4, Appli
	42	23	35.9	4044	12	US-10-014-927-18	Sequence 18, Appl
c	43	23	35.9	7470	9	US-10-073-961-397	Sequence 397, App
c	44	23	35.9	7470	9	US-10-073-961-398	Sequence 398, App
c	45	23	35.9	7470	10	US-09-764-887-397	Sequence 397, App

ALIGNMENTS

RESULT 1
US-09-783-590-762/c
; Sequence 762, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 762
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (44)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (52)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (62)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (106)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (123)

```
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (345)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (389)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (423)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (426)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (441)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (445)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-762

  Query Match             41.9%;  Score 26.8;  DB 10;  Length 445;
  Best Local Similarity   68.5%;  Pred. No. 2.6;
  Matches   37;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       9 AGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTAG 62
          ||||| | |||  |||||||  | ||||  | || | || ||  ||||  ||||
Db     232 AGTCCCCAGATTGGGCTTGAGCAATTCACATTCATAACCCTTGATACAAGGTAG 179


RESULT 2
US-09-918-995-30115/c
; Sequence 30115, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30115
;   LENGTH: 362
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(362)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30115

  Query Match             38.4%;  Score 24.6;  DB 9;  Length 362;
  Best Local Similarity   61.9%;  Pred. No. 16;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
          |||| |||  | | |||  | ||||   |  ||||||| |   | | | ||  ||| |||
Db     343 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 284

Qy      62 GCC 64
           | |
Db     283 GGC 281


RESULT 3
US-09-918-995-7625/c
; Sequence 7625, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7625
;   LENGTH: 394
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(394)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-7625

  Query Match             38.4%;  Score 24.6;  DB 9;  Length 394;
  Best Local Similarity   61.9%;  Pred. No. 16;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
          |||| |||  | | |||  | ||||   |  ||||||| |   | | | ||  ||| |||
Db     363 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 304

Qy      62 GCC 64
           | |
Db     303 GGC 301


RESULT 4
US-09-918-995-8119/c
; Sequence 8119, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8119
;   LENGTH: 425
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(425)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8119

  Query Match             38.4%;  Score 24.6;  DB 9;  Length 425;
  Best Local Similarity   61.9%;  Pred. No. 17;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
          |||| |||  | | |||  | ||||   |  ||||||| |   | | | ||  ||| |||
Db     224 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 165

Qy      62 GCC 64
           | |
Db     164 GGC 162


RESULT 5
US-09-918-995-31162/c
```

```
; Sequence 31162, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31162
;   LENGTH: 468
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(468)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31162

  Query Match             38.4%;  Score 24.6;  DB 9;  Length 468;
  Best Local Similarity   61.9%;  Pred. No. 17;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||  |  |||||||  |   | | | ||  ||| |||
Db     253 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 194

Qy      62 GCC 64
           | |
Db     193 GGC 191


RESULT 6
US-09-939-980-180
; Sequence 180, Application US/09939980
; Patent No. US20020082234A1
;   GENERAL INFORMATION:
;        APPLICANT: Black, Michael
;                   Burnham, Martin
;                   Hodgson, John
;                   Knowles, David
;                   Lonetto, Michael
;                   Nicholas, Richard
;                   Pratt, Julie
;                   Reichard, Richard
;                   Rosenberg, Martin
;                   Ward, Judith
;        TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,
;                            Polypeptides and Their Uses
;        NUMBER OF SEQUENCES: 534
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: SmithKline Beecham Corporation
;             STREET: 709 Swedeland Road
;             CITY: King of Prussia
;             STATE: PA
;             COUNTRY: USA
;             ZIP: 19406-0939
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/939,980
;             FILING DATE: 27-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/936,165
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Gimmi, Edward R
;             REGISTRATION NUMBER: 38,891
;             REFERENCE/DOCKET NUMBER: P50549
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 610-270-4478
;             TELEFAX: 610-270-5090
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 180:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 443 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: Genomic DNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 180:
US-09-939-980-180

  Query Match             38.1%;  Score 24.4;  DB 10;  Length 443;
  Best Local Similarity   73.8%;  Pred. No. 20;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      16 TGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACAT 57
           |||| ||| || ||| ||||||||  || || ||| |  | ||
Db     200 TGATTAGGTTTCAAAGGTTCATTTTCAAACACATCTAAATAT 241


RESULT 7
US-09-822-830A-561/c
; Sequence 561, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
;  APPLICANT: Genetics Institute, Inc.
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6402
;  CURRENT APPLICATION NUMBER: US/09/822,830A
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/195,604
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 631
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 561
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 589,681,785,788,790,810,815
;   OTHER INFORMATION: n=a,c,g, or t
US-09-822-830A-561

  Query Match             38.1%;  Score 24.4;  DB 10;  Length 826;
  Best Local Similarity   63.8%;  Pred. No. 25;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       4 AAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||||| ||   ||||| |  ||| |||  |  |||||||  |   | ||| | |  ||
Db     106 AAAAAAATCTCTTGATATGTATTGTAAATATAGTTTCCAGTCTATTGTTTGCCTTTTA 49


RESULT 8
US-09-938-842A-3526
; Sequence 3526, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3526
;   LENGTH: 1245
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-3526

  Query Match             38.1%;  Score 24.4;  DB 9;  Length 1245;
  Best Local Similarity   68.0%;  Pred. No. 29;
  Matches   34;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       5 AAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTA 54
           ||||| |||| |||  || |   || ||| || ||||| | |  || |||
Db    1011 AAAAACTCCGTTGAGTAGACAAAAAGAGTGCAATTCCAAAGCATTTCTTA 1060


RESULT 9
US-08-781-986A-551/c
; Sequence 551, Application US/08781986A
; Publication No. US20030054436A1
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Staphylococcus aureus Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  5255
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,986A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Bob
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PB248PP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  551:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4923 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-781-986A-551

  Query Match             38.1%;  Score 24.4;  DB 7;  Length 4923;
  Best Local Similarity   73.8%;  Pred. No. 48;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      16 TGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACAT 57
           |||| ||| || ||| |||||||| || || ||| |  | ||
Db    4116 TGATTAGGTTTCAAAGGTTCATTTTCAAACACATCTAAATAT 4075


RESULT 10
US-09-974-300-7252/c
; Sequence 7252, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7252
;   LENGTH: 1043
;   TYPE: DNA
;   ORGANISM: Bacillus clausii
US-09-974-300-7252

  Query Match             37.8%;  Score 24.2;  DB 10;  Length 1043;
  Best Local Similarity   66.0%;  Pred. No. 33;
  Matches   35;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTA 54
           | |||||  |||  ||| ||||| || |||   | || | || |||||  | |
Db     159 AGAAAAATCTCCAGTGAAAAGGCCTGCAAACGCCTTTACTAGCCCCATGATGA 107


RESULT 11
US-09-974-300-4354
; Sequence 4354, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4354
;   LENGTH: 1228
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis
US-09-974-300-4354

  Query Match             37.8%;  Score 24.2;  DB 10;  Length 1228;
  Best Local Similarity   66.0%;  Pred. No. 35;
  Matches   35;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTT 53
           | ||||||| ||||||||   |   |||||||||||| |       |  |||| ||
Db     176 CCAAAAAATGTCCGCTGCCGATCTTTGAAAAGTTCGTCGAAGAAGGCATTATT 228
```

```
RESULT 12
US-09-878-574-796/c
; Sequence 796, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 796
;   LENGTH: 380
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: LIB3028-046-Q1-B1-B10
US-09-878-574-796

  Query Match             37.2%;  Score 23.8;  DB 10;  Length 380;
  Best Local Similarity   62.7%;  Pred. No. 32;
  Matches   37;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           || ||||||  |    | |||     ||   ||||||    ||| ||||||||||| ||
Db     183 AATAAAAAGCACAACAAGAAGCTACTAATTTTTCATTAGGTGACTCATTTTTACATGGT 125


RESULT 13
US-09-815-242-6976
; Sequence 6976, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6976
;   LENGTH: 1371
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1371)
US-09-815-242-6976

  Query Match             37.2%;  Score 23.8;  DB 10;  Length 1371;
  Best Local Similarity   66.7%;  Pred. No. 51;
  Matches   34;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTT 52
           ||||||    ||| || ||| | ||  |||||  ||  || |||| || ||
Db    1051 AAAAAATCTACCGTTGGTAAAGGTTCTAAAGTAAATCACCTGACCTATGTT 1101


RESULT 14
US-10-003-806-10/c
; Sequence 10, Application US/10003806
; Patent No. US20020119929A1
; GENERAL INFORMATION:
;  APPLICANT: Bishop, Colin E.
;  APPLICANT:  Agoulnik, Alexander I.
;  APPLICANT:  Zhu, Qichao
;  TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
;  FILE REFERENCE: P02066US1/10024824
;  CURRENT APPLICATION NUMBER: US/10/003,806
;  CURRENT FILING DATE:  2001-11-02
;  PRIOR APPLICATION NUMBER: 60/245,872
;  PRIOR FILING DATE: 2000-11-03
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;   LENGTH: 173808
;   TYPE: DNA
;   ORGANISM: Mouse
US-10-003-806-10

  Query Match             36.9%;  Score 23.6;  DB 12;  Length 173808;
  Best Local Similarity   61.3%;  Pred. No. 3.4e+02;
  Matches   38;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           |||||| || || ||  | |     ||  |||     || ||||| | ||  | ||||| |
Db  116783 CAAAAACAACTCAGCCAAAAGAAGATGCAAAAGATGTTCCCAGAACAATGGTAACATCAT 116724

Qy      61 AG 62
           ||
Db  116723 AG 116722


RESULT 15
US-09-896-852-45/c
; Sequence 45, Application US/09896852
; Patent No. US20020025542A1
; GENERAL INFORMATION:
;  APPLICANT: Abbott Laboratories
;  APPLICANT:  Maine, Gregory T.
;  APPLICANT:  Hunt, Jeffery C.
;  APPLICANT:  Brojanac, Susan
;  APPLICANT:  Jhy-Tsing Sheu, Michael
;  APPLICANT:  Chovan, Linda E.
;  APPLICANT:  Tyner, Joan D.
;  APPLICANT:  Howard, Lawrence V.
;  TITLE OF INVENTION: ANTIGEN COCKTAILS, P35, AND USES THEREOF
;  FILE REFERENCE: 6361.US.D1
;  CURRENT APPLICATION NUMBER: US/09/896,852
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 09/086,503
;  PRIOR FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 4451
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
```

US-09-896-852-45

Query Match 36.6%; Score 23.4; DB 10; Length 4451;
Best Local Similarity 73.2%; Pred. No. 1.1e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCA 42
           ||||||||| ||||| || ||||  | . ||| |  || |||
Db    1686 AAAAAAAAGCCCGCTCATTAGGCGGGCTAAGCTTCTTGCCA 1646
```

Search completed: May 5, 2003, 18:05:48
Job time : 48.3626 secs

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:53:33 ; Search time 390.505 Seconds
(without alignments)
2654.281 Million cell updates/sec

Title: US-09-990-099-23
Perfect score: 64
Sequence: 1 caaaaaaaagtccgctgata.....cccatttttacatcgtagcc 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	38	59.4	863	17	AF094932	AF094932 AF094932
	2	28.4	44.4	700	9	AJ425333	AJ425333 AJ425333
	3	28.4	44.4	750	9	AL564193	AL564193 AL564193
c	4	28.2	44.1	379	10	AV820230	AV820230 AV820230
c	5	28	43.8	215	10	BB014998	BB014998 BB014998
	6	27.8	43.4	518	10	AV699330	AV699330 AV699330
c	7	27.8	43.4	580	13	BJ529330	BJ529330 BJ529330
c	8	27.8	43.4	620	10	AV662150	AV662150 AV662150
	9	26.8	41.9	384	14	T51247	T51247 yb03e12.s1
c	10	26.8	41.9	680	9	AI061603	AI061603 HA0370 Hu
	11	26.8	41.9	740	17	AQ954300	AQ954300 nbeb0066N
	12	26.6	41.6	742	17	AQ867715	AQ867715 nbeb0034K
	13	26.6	41.6	934	12	BG443824	BG443824 GA__Ea002
	14	26.4	41.2	288	9	AA343029	AA343029 EST48667
	15	26.4	41.2	335	14	T67877	T67877 yc39f02.s1
	16	26.4	41.2	355	14	T67504	T67504 yc26g08.s1
c	17	26.4	41.2	559	13	BI775172	BI775172 467563 MA
c	18	26.4	41.2	559	13	BI775175	BI775175 467569 MA
c	19	26.4	41.2	569	9	AA046980	AA046980 zf13c09.r
c	20	26.4	41.2	587	13	BM288511	BM288511 529796 MA
c	21	26.4	41.2	1005	17	CNS00ZLA	AL097816 Drosophil
	22	26.2	40.9	290	14	T41012	T41012 ya15g08.s3
	23	26.2	40.9	307	13	BM158733	BM158733 NXLV_039_
	24	26.2	40.9	379	14	N58056	N58056 yv63h01.s1
c	25	26.2	40.9	407	10	AV649451	AV649451 AV649451
c	26	26.2	40.9	440	9	AI285962	AI285962 qh97g07.x
	27	26.2	40.9	477	9	AI185952	AI185952 qe29c10.s
	28	26.2	40.9	501	9	AA723197	AA723197 zg88h05.s
c	29	26.2	40.9	504	9	AI065114	AI065114 HA0951 Hu
c	30	26.2	40.9	553	12	BG565748	BG565748 602589084
	31	26.2	40.9	606	12	BF241003	BF241003 601880713
c	32	26.2	40.9	633	10	AV699943	AV699943 AV699943
c	33	26.2	40.9	639	10	AV682396	AV682396 AV682396
c	34	26.2	40.9	755	12	BG564841	BG564841 602584082
	35	26	40.6	282	14	T74865	T74865 yc58a03.s1
	36	26	40.6	415	14	H54054	H54054 yq88b09.s1
c	37	26	40.6	435	12	BG044593	BG044593 saa30b11.
c	38	26	40.6	604	9	AI541136	AI541136 pec1.2-11
	39	26	40.6	656	17	BH111596	BH111596 RPCI-24-2
	40	26	40.6	673	17	AG173379	AG173379 Pan trog1
	41	26	40.6	1097	12	BE908674	BE908674 601499512
	42	26	40.6	1108	12	BE908541	BE908541 601499610
	43	26	40.6	1398	14	BQ952447	BQ952447 AGENCOURT
	44	25.8	40.3	222	14	T57263	T57263 yb33g07.s1
	45	25.8	40.3	230	9	AA342380	AA342380 EST47929

ALIGNMENTS

RESULT 1
AF094932/c
LOCUS AF094932 863 bp DNA linear GSS 29-AUG-2000
DEFINITION AF094932 Salmonella typhimurium LT2, Lambda DASH II Salmonella typhimurium genomic clone 1335-T7, DNA sequence.
ACCESSION AF094932
VERSION AF094932.1 GI:4322774
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 863)
AUTHORS Wong,R.M.Y. and McClelland,M.
TITLE End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones, Li-Cor
JOURNAL Unpublished (1999)
COMMENT Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mcclelland@lifsci.sdsu.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .863
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/clone="1335-T7"

```
                     /clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
                     /note="Vector: Lambda DASH II; sequenced using Li-Cor
                     sequencer"
BASE COUNT       209 a    213 c    205 g    236 t
ORIGIN

  Query Match             59.4%;  Score 38;  DB 17;  Length 863;
  Best Local Similarity   75.8%;  Pred. No. 0.0014;
  Matches   47;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           |||||||||| | | | | ||  |||||||||| | | | |||||||||||  ||   |||
Db     854 CAAAAAAAAGCCGGATTAAAATACTTGAAAAGTACGTATTCAGACCCATTTATAAGACGT 795

Qy      61 AG 62
           ||
Db     794 AG 793


RESULT 2
AJ425333
LOCUS       AJ425333                700 bp    mRNA    linear   EST 31-JUL-2002
DEFINITION  AJ425333 Salmo salar spleen adult Salmo salar cDNA clone s10C11,
            mRNA sequence.
ACCESSION   AJ425333
VERSION     AJ425333.1  GI:22033501
KEYWORDS    EST.
SOURCE      Atlantic salmon.
  ORGANISM  Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Cairney,M. and Taggart,J.
  TITLE     Gene sequences in atlantic salmon (Salmo salar)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Taggart JB
            Institute of Aquaculture
            University of Stirling
            Airthrey Road, FK9 4LA, United Kingdom.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Salmo salar"
                     /db_xref="taxon:8030"
                     /clone="s10C11"
                     /clone_lib="Salmo salar spleen adult"
                     /tissue_type="spleen"
                     /dev_stage="adult"
                     /note="Vector: Stratagene lambda Uni-Zap; country=United
                     Kingdom"
BASE COUNT       193 a    162 c    168 g    177 t
ORIGIN

  Query Match             44.4%;  Score 28.4;  DB 9;  Length 700;
  Best Local Similarity   70.4%;  Pred. No. 5.2;
  Matches   38;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTA 54
           ||| ||||||||  | |  | ||||  || |||| |||||  ||||||  || |
Db     110 CAACAAAAAGTCTACAGCGATGGCTGCAACAGTTGATTTCAGGACCCACGTTGA 163


RESULT 3
AL564193
LOCUS       AL564193                750 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL564193 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM003YO11 3
            prime, mRNA sequence.
ACCESSION   AL564193
VERSION     AL564193.1  GI:12914354
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DM003YO11"
                     /clone_lib="LTI_NFL001_NBC4"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT       197 a    148 c    150 g    231 t     24 others
ORIGIN

  Query Match             44.4%;  Score 28.4;  DB 9;  Length 750;
  Best Local Similarity   65.1%;  Pred. No. 5.3;
  Matches   41;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           ||||||||  | | ||   |  |||  ||| ||||||| ||  | | | ||  ||| |||
Db      74 AAAAAAAAAACAGATGNATAAACTTTTAAACTTCATTTTCATTCTCTTGTTCTCATGGTA 133

Qy      62 GCC 64
           | |
Db     134 GGC 136


RESULT 4
AV820230/c
LOCUS       AV820230                379 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV820230 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-09-C19 3',
            mRNA sequence.
ACCESSION   AV820230
VERSION     AV820230.1  GI:19862189
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
```

and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further details.

FEATURES Location/Qualifiers
source 1. .379
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="RAFL11-09-C19"
/clone_lib="RAFL11"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/note="Site_1: BamHI; Site_2: SalI; subjected to various treatments (dehydration, cold, high salt, ABA, heat and UV). Dark-grown plants"

BASE COUNT 132 a 81 c 64 g 102 t
ORIGIN

Query Match 44.1%; Score 28.2; DB 10; Length 379;
Best Local Similarity 68.4%; Pred. No. 5.5;
Matches 39; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACAT 57
           ||||| |||  |||||||| |||  || |||||| ||  |   ||   ||||  |||
Db     197 CAAAACAAACCCCGCTGATGAGGAATGCAAAGTTTATGACTGAACAACTTTTATCAT 141
```

RESULT 5
BB014998/c
LOCUS BB014998 215 bp mRNA linear EST 22-JUN-2000
DEFINITION BB014998 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4930487D23 3', mRNA sequence.
ACCESSION BB014998
VERSION BB014998.1 GI:8138189
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 215)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .215
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4930487D23"
/clone_lib="RIKEN full-length enriched, adult male testis (DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 50 a 59 c 50 g 56 t
ORIGIN

Query Match 43.8%; Score 28; DB 10; Length 215;
Best Local Similarity 66.7%; Pred. No. 5.9;
Matches 40; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGT 60
           ||| |||||| | | || |   |  || ||||| | | ||||||| |||  ||  ||| |
Db      90 CAATAAAAAGGCTGTTGGTCTTGAATGCAAAGTCCTTCTCCAGACTCATGCTTGAATCCT 31
```

RESULT 6
AV699330
LOCUS AV699330 518 bp mRNA linear EST 16-JAN-2002
DEFINITION AV699330 GKC Homo sapiens cDNA clone GKCGEC01 3', mRNA sequence.
ACCESSION AV699330
VERSION AV699330.1 GI:10301301
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 518)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE 21625106
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)

```
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
  FEATURES             Location/Qualifiers
       source          1. .518
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="GKCGEC01"
                       /clone_lib="GKC"
                       /tissue_type="hepatocellular carcinoma"
                       /dev_stage="Adult"
                       /lab_host="SOLR"
                       /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                       XhoI"
  BASE COUNT       148 a      97 c     122 g     150 t       1 others
  ORIGIN

    Query Match             43.4%;  Score 27.8;  DB 10;  Length 518;
    Best Local Similarity   65.1%;  Pred. No. 8.3;
    Matches   41;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

  Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
             |||| |||  |||||||  | ||||   |  ||||||| |   | | | ||  ||| |||
  Db      81 AAAAGAAAAACCGCTGAATAAGCTTTGGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 140

  Qy      62 GCC 64
             | |
  Db     141 GGC 143


  RESULT 7
  BJ529330/c
  LOCUS       BJ529330                  580 bp    mRNA    linear   EST 09-AUG-2002
  DEFINITION  BJ529330 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB035I05 5',
              mRNA sequence.
  ACCESSION   BJ529330
  VERSION     BJ529330.1  GI:22188142
  KEYWORDS    EST.
  SOURCE      Japanese medaka.
    ORGANISM  Oryzias latipes
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
              Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
  REFERENCE   1  (bases 1 to 580)
    AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
    TITLE     Medaka EST Project in Takeda's lab
    JOURNAL   Unpublished (2001)
  COMMENT     Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
  FEATURES             Location/Qualifiers
       source          1. .580
                       /organism="Oryzias latipes"
                       /strain="Hd-rR"
                       /db_xref="taxon:8090"
                       /clone="MF01SSB035I05"
                       /clone_lib="MF01SSB cDNA"
                       /sex="mixture of female and male"
                       /tissue_type="whole embryo"
                       /dev_stage="segmentation stage 20 - 25"
  BASE COUNT       168 a     100 c      97 g     209 t       6 others
  ORIGIN

    Query Match             43.4%;  Score 27.8;  DB 13;  Length 580;
    Best Local Similarity   69.1%;  Pred. No. 8.5;
    Matches   38;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

  Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACA 56
             |||||||| ||  |||  |  |||| ||||||||||||  |  ||   || || |
  Db     251 AAAAAAAAATCTCCTGCAACAGCTTTAAAAGTTCATTTGTAAGCCGGATTATAAA 197


  RESULT 8
  AV662150/c
  LOCUS       AV662150                  620 bp    mRNA    linear   EST 16-JAN-2002
  DEFINITION  AV662150 GLD Homo sapiens cDNA clone GLDAAA02 3', mRNA sequence.
  ACCESSION   AV662150
  VERSION     AV662150.1  GI:9883164
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 620)
    AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
              Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
              Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
              Hu,G., Gu,J., Chen,Z. and Han,Z.
    TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
              by comparing gene expression profiles of hepatocellular carcinoma
              with those of corresponding noncancerous liver
    JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
    MEDLINE   21625106
  COMMENT     Contact: Zeguang Han
              Chinese National Human Genome Center at Shanghai
              351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
              201203, P. R. China
              Tel: 86-21-50801919(ex.45)
              Fax: 86-21-50801922
              Email: hanzg@chgc.sh.cn
              This clone is available at CHGC in Shanghai.
  FEATURES             Location/Qualifiers
       source          1. .620
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="GLDAAA02"
                       /clone_lib="GLD"
                       /tissue_type="corresponding non cancerous liver tissue"
                       /dev_stage="Adult"
                       /lab_host="SOLR"
                       /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                       XhoI"
  BASE COUNT       227 a     103 c     114 g     176 t
  ORIGIN

    Query Match             43.4%;  Score 27.8;  DB 10;  Length 620;
    Best Local Similarity   65.1%;  Pred. No. 8.6;
    Matches   41;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

  Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
             |||| |||  | | |||  | ||||   |  ||||||| | | | | ||  ||| |||
  Db     363 AAAAGAAAAACAGATGAATAAGCTTTTGATCTTCATTTTCTGTCTCTTATTCTCATGGTA 304

  Qy      62 GCC 64
             |||
  Db     303 GCC 301


  RESULT 9
  T51247
  LOCUS       T51247                    384 bp    mRNA    linear   EST 06-FEB-1995
  DEFINITION  yb03e12.s1 Stratagene placenta (#937225) Homo sapiens cDNA clone
              IMAGE:70126 3' similar to similar to gb:A06977 SERUM ALBUMIN
              PRECURSOR (HUMAN), mRNA sequence.
  ACCESSION   T51247
  VERSION     T51247.1  GI:653107
  KEYWORDS    EST.
  SOURCE      human.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 671
            High qality sequence stops: 306 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 671    Std Error: 0.00
            Seq primer: -21m13
            High quality sequence stop: 306.
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Homo sapiens"
                     /db_xref="GDB:491023"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:70126"
                     /clone_lib="Stratagene placenta (#937225)"
                     /sex="male"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: placenta; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Caucasian. Average insert size: 1.2 kb; Uni-ZAP
                     XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      107 a     70 c     72 g    127 t      8 others
ORIGIN

  Query Match             41.9%;  Score 26.8;  DB 14;  Length 384;
  Best Local Similarity   63.5%;  Pred. No. 19;
  Matches   40;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||   || ||||||| |   | | | ||  ||| |||
Db     102 AAAAGAAAAACAGATGAATAAGCTTTTGAACTTCATTTNCTTTCTCTTATTCTCATGGTA 161

Qy      62 GCC 64
           | |
Db     162 GGC 164


RESULT 10
AI061603/c
LOCUS       AI061603                680 bp    mRNA    linear   EST 11-NOV-1999
DEFINITION  HA0370 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   AI061603
VERSION     AI061603.1  GI:6358900
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
            and He,F.
  TITLE     Expression profile analysis of a human fetal liver cDNA library
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Yongtao Yu
            Department of Hematology .
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Tel: 0086-10-68159479
            Fax: 0086-10-68214653
            Email: yyt48@yahoo.com.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human fetal liver cDNA library"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="MC1061/P3"
                     /note="Vector: pCDNA1"
BASE COUNT      227 a    140 c    143 g    168 t      2 others
ORIGIN

  Query Match             41.9%;  Score 26.8;  DB 9;  Length 680;
  Best Local Similarity   63.5%;  Pred. No. 21;
  Matches   40;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||   || ||||||| |   | | | ||  ||| |||
Db     628 AAAAGAAAAACAGATGAATAAGCTTNAGATCTTCATTTTCTTTCTCTTATTCTCATGGTA 569

Qy      62 GCC 64
           | |
Db     568 GGC 566


RESULT 11
AQ954300
LOCUS       AQ954300                740 bp    DNA     linear   GSS 27-JAN-2000
DEFINITION  nbeb0066N06f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone OSJNBb0066N06f, DNA sequence.
ACCESSION   AQ954300
VERSION     AQ954300.1  GI:6777565
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GTAAAACGACGGCCAGTG
            Class: BAC ends
            High quality sequence start: 56
            High quality sequence stop: 312.
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="OSJNBb0066N06f"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
```

haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximatley 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 224 a 157 c 123 g 235 t 1 others
ORIGIN

Query Match 41.9%; Score 26.8; DB 17; Length 740;
Best Local Similarity 68.5%; Pred. No. 21;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy       1 CAAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTA 54
           ||||  ||| | |  | ||| ||  |  || | | ||||||| ||||||| |||
Db     639 CAAAGGAAATTTCCTTTATACGGGATCTAAGGATAATTTCCAAACCCATTCTTA 692
```

RESULT 12
AQ867715
LOCUS AQ867715 742 bp DNA linear GSS 03-NOV-1999
DEFINITION nbeb0034K11r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic clone nbeb0034K11r, DNA sequence.
ACCESSION AQ867715
VERSION AQ867715.1 GI:6218276
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 742)
AUTHORS Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAAACAGCTATGACCATG
Class: BAC ends
High quality sequence start: 23
High quality sequence stop: 460.
FEATURES Location/Qualifiers
source 1. .742
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbeb0034K11r"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI; Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximatley 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 249 a 137 c 134 g 221 t 1 others
ORIGIN

Query Match 41.6%; Score 26.6; DB 17; Length 742;
Best Local Similarity 70.0%; Pred. No. 25;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTT 51
           | | | |||||| |||||| ||  || ||| |  | || ||||| | |||
Db     210 ACANATAAGTCCTCTGATACGGTGTGCAAACTGAACTTTCAGACACTTTT 259
```

RESULT 13
BG443824
LOCUS BG443824 934 bp mRNA linear EST 15-MAR-2001
DEFINITION GA__Ea0022E01f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA__Ea0022E01f, mRNA sequence.
ACCESSION BG443824
VERSION BG443824.1 GI:13353476
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 934)
AUTHORS Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 665.
FEATURES Location/Qualifiers
source 1. .934
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA__Ea0022E01f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 276 a 212 c 239 g 206 t 1 others
ORIGIN

Query Match 41.6%; Score 26.6; DB 12; Length 934;
Best Local Similarity 78.0%; Pred. No. 26;
Matches 32; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy      23 GCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTAGC 63
```

```
                 ||||| | || ||||||| |   |||||||| || ||||||
Db     383 GCTTGCAGAGGTCATTTCGAAGACCATTTTTTCACCGTAGC 423

RESULT 14
AA343029
LOCUS       AA343029                 288 bp    mRNA    linear   EST 21-APR-1997
DEFINITION  EST48667 Fetal spleen Homo sapiens cDNA 3' end similar to similar
            to albumin, mRNA sequence.
ACCESSION   AA343029
VERSION     AA343029.1  GI:1995265
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 288)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):144514"
                     /db_xref="taxon:9606"
                     /clone_lib="Fetal spleen"
                     /dev_stage="fetus"
                     /note="Organ: spleen; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       85 a     53 c     52 g     91 t      7 others
ORIGIN

  Query Match             41.2%;  Score 26.4;  DB 9;  Length 288;
  Best Local Similarity   61.9%;  Pred. No. 25;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||   |  |||||||  |   | | | ||  ||| |||
Db     102 AAAAGAAAAACAGNTGANTAAGCTTTTGANCTTCATTTTCTTTCTCTTATTCTCATGGTA 161

Qy      62 GCC 64
           | |
Db     162 GGC 164


RESULT 15
T67877
LOCUS       T67877                   335 bp    mRNA    linear   EST 22-FEB-1995
DEFINITION  yc39f02.s1 Stratagene liver (#937224) Homo sapiens cDNA clone
            IMAGE:83067 3' similar to gb:A06977 SERUM ALBUMIN PRECURSOR (HUMAN
            );; mRNA sequence.
ACCESSION   T67877
VERSION     T67877.1  GI:679025
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High qality sequence stops: 234
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -21m13
            High quality sequence stop: 234.
FEATURES             Location/Qualifiers
     source          1. .335
                     /organism="Homo sapiens"
                     /db_xref="GDB:500124"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:83067"
                     /clone_lib="Stratagene liver (#937224)"
                     /sex="male"
                     /dev_stage="49 years old"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: liver; Vector: pBluescript SK; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. Hepatectomy from normal male caucasian. Average insert
                     size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       84 a     73 c     63 g     99 t      16 others
ORIGIN

  Query Match             41.2%;  Score 26.4;  DB 14;  Length 335;
  Best Local Similarity   61.9%;  Pred. No. 26;
  Matches   39;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 AAAAAAAAGTCCGCTGATAAGGCTTGAAAAGTTCATTTCCAGACCCATTTTTACATCGTA 61
           |||| |||  | | |||  | ||||   |  |||||||  |   | | | ||  ||| |||
Db      29 AAAAGAAAAACAGATGAATAAGCTTTTGANCTTCATTTNCTTNCTCTTATTCTCATGGTA 88

Qy      62 GCC 64
           | |
Db      89 GGC 91


Search completed: May  5, 2003, 18:41:09
Job time : 395.505 secs
```